

FEATURE SELECTION
FOR THE
IDENTIFICATION OF BAUDED SIGNALS

Thomas Arthur Reglein

JUL 14 1964 X LIBRARY

DATE _____

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_X f d\mu$$

NAVAL POSTGRADUATE SCHOOL

Monterey, California



THESIS

FEATURE SELECTION
FOR THE
IDENTIFICATION OF BAUDED SIGNALS

by

Thomas Arthur Reglein

March 1975

Thesis Advisor:

S. Jauregui

Approved for public release; distribution unlimited.

T167565

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Feature Selection for the Identification of Bauded Signals		5. TYPE OF REPORT & PERIOD COVERED Engineer's Thesis; March 1975
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas Arthur Reglein		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE March 1975
		13. NUMBER OF PAGES 154
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Pattern Recognition Signal Identification Parameter Encoder Feature Selection		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This paper investigates some of the difficulties commonly encountered in the application of pattern recognition techniques to the identification of non-cooperative signals. Practical problems of feature selection and analysis are explored for a data base of actual banded signal transition times. Analysis of a set of features based on the raster display of these transition times suggests that the use of		

(20. ABSTRACT Continued)

fast Fourier transform techniques may yield useful identifying parameters not exploitable by the current process of visual comparison of signal raster patterns.

Feature Selection
for the
Identification of Bauded Signals

by

Thomas Arthur Reglein
Lieutenant, United States Navy
B.S., Iowa State University, 1967

Submitted in partial fulfillment of the
requirements for the degree of

ELECTRICAL ENGINEER

from the
NAVAL POSTGRADUATE SCHOOL
March 1975

ABSTRACT

This paper investigates some of the difficulties commonly encountered in the application of pattern recognition techniques to the identification of non-cooperative signals. Practical problems of feature selection and analysis are explored for a data base of actual banded signal transition times. Analysis of a set of features based on the raster display of these transition times suggests that the use of fast Fourier transform techniques may yield useful identifying parameters not exploitable by the current process of visual comparison of signal raster patterns.

TABLE OF CONTENTS

I.	INTRODUCTION -----	11
II.	BAYESIAN PATTERN CLASSIFICATION -----	13
	A. PATTERN LEARNING -----	15
	B. DECISION RULES -----	16
III.	PRACTICAL CONSIDERATIONS IN IDENTIFICATION OF NON-COOPERATIVE SIGNALS -----	20
	A. SMALL SAMPLE SIZE -----	21
	B. INDETERMINANCY OF CLASSES -----	24
	C. INACCURATE GROUND TRUTH -----	27
IV.	FEATURE SELECTION FOR WAVEFORM IDENTIFICATION ---	29
V.	PARAMETER MAPPING OF RASTER SCAN DISPLAY -----	33
	A. WAVEFORM GENERATION PROCESS -----	34
	B. RASTER SCAN DISPLAY -----	24
	C. RASTER SCAN DESCRIPTIVE PARAMETERS -----	46
	1. Mean Bias and Transients -----	48
	2. Signal Variance and Intrinsic Variance --	52
	3. Polynomial Mean Square Fit -----	55
	4. Fourier Coefficients -----	56
	5. Laguerre Polynomials -----	60
	D. COMPUTER PROGRAM DESCRIPTIONS -----	73
	1. Measurement Process -----	74
	2. Statistical Programs -----	80
	E. ANALYSIS OF PARAMETERS -----	83
	1. Single Feature Separability -----	86
	2. Multiple Feature Analysis -----	91

VI. SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS -----	104
APPENDIX A: METHODS OF PROBABILITY ESTIMATION -----	107
APPENDIX B: CLASSIFICATION ALGORITHMS -----	113
APPENDIX C: METHODS OF FEATURE SELECTION -----	117
PARAMETER HISTOGRAMS-----	130
COMPUTER PROGRAMS -----	136
LIST OF REFERENCES -----	152
INITIAL DISTRIBUTION LIST -----	154

LIST OF FIGURES

Figure 1.	Misclassification Rates Due to Suboptimal Decision Boundaries -----	18
Figure 2.	Two-Dimensional Guard Zone Clustering -----	26
Figure 3.	Raster Signal Generation Process -----	35
Figure 4.	Relation of Phase Perturbation to Raster Display -----	38
Figure 5.	Raster Display of Constant Period Events t_n -----	41
Figure 6.	Raster Display of Bauded Signal Bias -----	43
Figure 7.	Quantization and Aliasing Effects in Raster Display Period = 18.6 x Quantization Interval -----	44
Figure 8.	Quantization and Aliasing Effects in Raster Display Period = 20.0 x Quantization Interval -----	45
Figure 9.	Minimal Phase Difference Raster Display ----	47
Figure 10.	Relation of Exponential Transient to Raster Display -----	51
Figure 11.	Effect of Transient on Normal Raster Display -----	53
Figures 12-22.	Performance of Polynomial Mean Square Fit, Interpolation Procedure and Fourier Magnitude Measurement on Five Simulated Signals -----	61-71
Figure 23.	Process Flow Interface and Raster Display Programs FACE and RASF -----	76
Figure 24.	Process Flow Measurement and Display Program FOURD -----	77

Figure 25.	Flow Diagram, Interpolation and Measurement Sub Routine RESDU -----	78
Figure 26.	Flow Diagram, Minimal Phase Difference Algorithm -----	79
Figure 27.	Feature Separation Quality vs. Frequency, 59 Signals, 7 classes -----	89
Figure 28.	Feature Separation Quality vs. Frequency, 30 Signals, 4 classes -----	90
Figure 29-30.	Scattergram of Class Membership, Signal Duration vs. Nominal Period Raster ---	93-94
Figure 31-32.	Scattergram of Class Membership, Squared Magnitude of 18 Hz Frequency Component vs. Nominal Raster Period -----	95-96
Figures 33-38.	Feature Histograms Showing Class Membership for Six Parameters Selected in Search Routine -----	130- 135

LIST OF TABLES

Table I	File Structure for Disc File PARF -----	75
Table II	File Structure for Disc File PAR -----	81
Table III	File Structure for Disc File MASTF -----	82
Table IV	Master File Parameter Number Identification -----	84
Table V	Training Set Signal Distribution -----	85
Table VI	Distance Ratios for Parameters 901-950 -----	88
Table VII	Parameters Ranked by Distance Ratio -----	92
Table VIII	Classifier Selection of Four Classes Nearest to Training Set Samples -----	98
Table IX	Confusion Matrix for Classifier Using Parameters 902, 903, 918, 926, 927, 931 -----	100
Table X	Confusion Matrix for Classifier Using Parameter 901 Only -----	101
Table XI	Confusion Matrix for Classifier Using Parameters 901, 903, 918, 926, 931 -----	103

ACKNOWLEDGEMENTS

I wish to express my deep appreciation to Mr. Byron Danzer and Electromagnetic Systems Laboratory for providing encouragement, technical support and facilities essential to this project. I am also grateful to CDR Henry Orjuela and LCDR Russ Shields for expressing active interest on the part of Naval Security Group Headquarters and PME-107. Above all, I must thank my wife Kay for her consideration and loyal assistance throughout the course of this work.

I. INTRODUCTION

The manual measurement and analysis of signal parameters leading to identification of the signal source can require more time and manpower than results justify. Such a process may be made less costly by the application of computer technology to measure the traditional parameters in a manner relatively free of individual bias and to aid in the analysis and identification procedures. An additional result of the computer application is the capability to calculate parameters unmeasurable by manual techniques. This thesis investigates the process by which such parameters might be obtained and used in an existing pattern recognition scheme. The discussion progresses from the general principles of Bayesian classification to the problem of finding features which are useful in classifying signals from a specific data base.

Section II of this thesis deals with the theory of pattern recognition which underlies most automated classification schemes. Section III discusses some of the limitation imposed on this theory in a practical signal identification problem. Section IV is devoted to the problems which arise in the attempt to choose a set of features useful and meaningful to the identification process. Section V addresses the particular problems of choosing a set of useful features based on the raster scan display of banded

signal transition times. Included in this section are the motivation for the initial choice of parameters and the techniques used for both measurement and analysis. The conclusions reached by this investigator, as well as those areas deserving further study, are summarized in Section VI.

II. BAYESIAN PATTERN CLASSIFICATION

The process of mathematical pattern recognition formalizes the methods by which the measured parameters or features of a group of samples may be used to identify those samples as belonging to one of several general classes. Such a process may be considered to be composed of two parts. The first part is a learning procedure in which the general "pattern" of each of the various classes is described in terms of the similarities and differences in the sample features. The second part is a decision process in which a feature-based algorithm is developed from the pattern description and applied to the identification or classification of new samples. Both parts of the process are based on the theory of statistical inference and lead to a probabilistic interpretation of the class membership. Closely related to such an interpretation is the Bayesian model of decision theory.

In the Bayesian model, the first step of a pattern recognition scheme, the learning process, is that technique by which a-priori class-conditional feature probabilities are determined. The K_i -individual d -dimensional feature vectors $\underline{x}_1^{(i)}, \underline{x}_2^{(i)}, \dots, \underline{x}_{K_i}^{(i)}$ of the m distinct classes C_i are used to estimate m class conditional probability density functions on the feature space

$$p(\underline{x}|C_i) \approx \hat{p}(\underline{x}|\underline{x}_1^{(i)} \dots \underline{x}_{K_i}^{(i)}) \quad (\text{II-1})$$

Knowing these distributions for each class and the overall class probabilities, $p(C_i)$, one may obtain the feature conditional class probabilities:

$$\begin{aligned} p(C_i|\underline{x}) &= \frac{p(C_i)p(\underline{x}|C_i)}{p(\underline{x})} \\ &= \frac{p(C_i)p(\underline{x}|C_i)}{\sum_{i=1}^m p(C_i)p(\underline{x}|C_i)} \end{aligned} \quad (\text{II-2})$$

Assuming an equal cost of misclassification, the Bayesian decision rule chooses that class which maximizes the feature conditional probabilities of Equation II-2. That is, if

$$p(C_i|\underline{x}) \geq p(C_j|\underline{x}) \quad \text{for all } j \quad (\text{II-3})$$

the sample characterized by the feature vector \underline{x} is assigned to class C_i .

The above technique generates a series of decision regions in the d-dimensional feature space. Once the a-priori probabilities are determined, the class selection is simply an m-valued function of the feature vector. This is essentially the goal of the pattern recognition process.

The probabilistic Bayesian decision process discussed above is mathematically elegant and optimal in the sense

that if the a-priori probabilities are known, it minimizes the expected risk of misclassification. Unfortunately, its implementation suffers major difficulties in both the learning and the decision phases of a statistical pattern recognition process.

A. PATTERN LEARNING

In the learning portion of a pattern recognition problem, one normally deals with a finite number of samples which constitute a training set. These samples are the basis from which the investigator is obliged to construct the necessary a-priori class-conditional probability distribution functions. Thus the well-defined probability structure assumed by Bayes' decision rule is in reality a statistical inference based on the data obtained from the training set and the investigator's best guess of the functional form of this data. Dependent on how much of the general form of the probability density function is presupposed, the estimation techniques are characterized as parametric or non-parametric.

Typically, parametric estimation assumes that the features of the training samples follow one of the well-known probability distributions. The samples of the training set are used to estimate only a few essential parameters of this distribution. For example, one might use the training samples to estimate the means and variances for an assumed multivariate normal distribution. The non-parametric technique on the other hand seeks to estimate directly the

probability density function as a superposition of "potential function" terms arising from each of the training set samples.

The accuracy of the inferred density function is strongly dependent on the type of estimation technique used, and may be severely limited by the size of the training set. The applicability and limitations of both non-parametric and parametric estimation are discussed further in Appendix A. Regardless of the type of estimation, the process involved is an example of supervised learning. It presupposes a set of correctly labeled samples from which the estimates may be constructed. Information external to the features being measured has been used to identify the members of each class.

B. DECISION RULES

The previous discussion of the Bayesian approach to signal classification already suggests one method of obtaining a decision rule. For each new (unclassified) feature vector use the a-priori probability inferred from the training set to calculate the feature-conditional probability of its membership in each class; then choose that class for which this probability is greatest.

Such a maximal likelihood criteria is intuitively satisfying and can be formulated to lead to classifications which are optimal in the sense of minimizing the average risk of misclassification. Unfortunately a naive implementation of such a technique leads to computational inefficiencies

which often render the process impractical in terms of memory and time constraints.

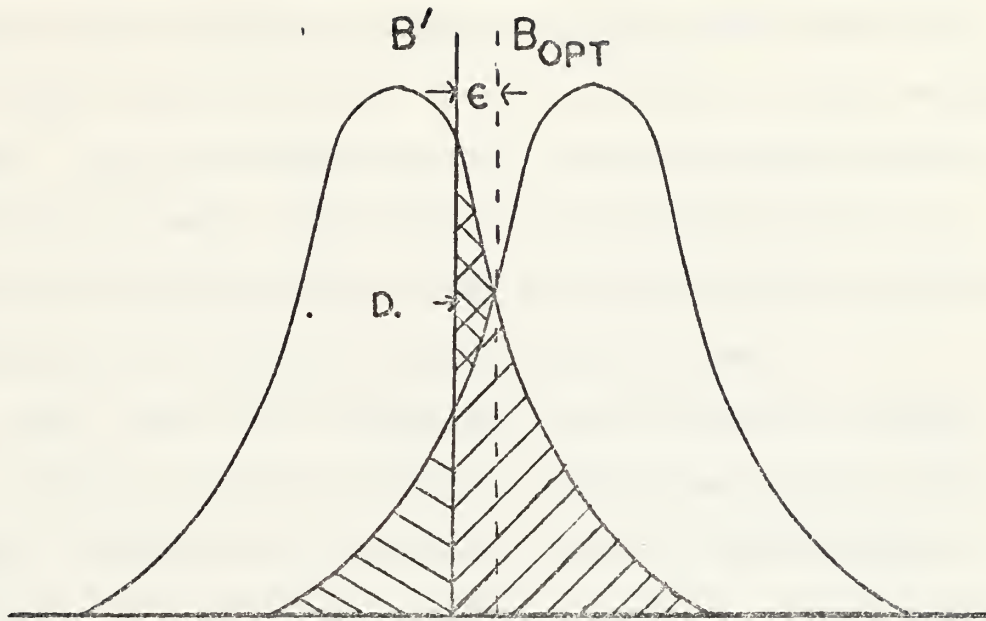
More frequently the decision rule chosen in a practical pattern recognition problem is suboptimal in the Bayesian sense, providing a method of partitioning the feature-space of the samples without resorting to the calculation of the maximal likelihood criteria. These suboptimal schemes use simplifying assumptions regarding the nature of the underlying class-conditional probability distributions to arrive at class boundaries which are relatively simple functions of the features themselves. If there is little overlap in the various class-conditional probability density functions, the increase in the expected misclassification rate as a result of the suboptimal decision is minimal.

For example, consider the case of the one-dimensional distributions for the two equiprobable classes shown in Figures 1a and 1b. Suppose that some suboptimal decision rule results in decision boundary

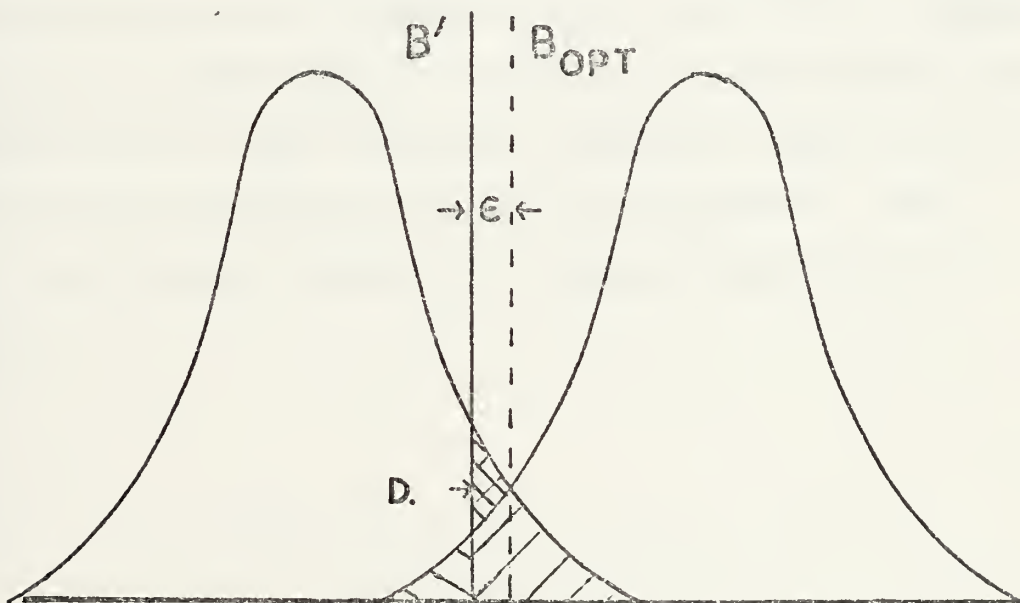
$$B' = B_{\text{opt}} - \epsilon \quad (\text{II-4})$$

The shaded areas of Figures 1a and 1b indicate the resultant misclassification rates.

If there is substantial separation of the class distributions, the differential error in classification as a result of boundary misplacement, area D, is negligible. In general,



(a) Large Overlap



(b) Small Overlap

FIGURE 1. Misclassification Rates Due to Suboptimal Decision Boundaries

with decreasing overlap of unimodal class distributions, not only does the total misclassification error decrease, but so does its sensitivity to the placement of the decision boundary. As a practical matter, the distinction between a five percent and a two percent misclassification rate seems unnecessarily fine if the probability distributions were originally inferred from ten samples each.

In many cases the sub-optimal classification schemes follow rather naturally from the method used to infer the a-priori distributions. Unimodal feature distributions give rise to classification schemes based on a test sample's minimal "distance" from estimated class means. More complex feature distributions involve the use of a distance measure in the feature space to establish the K training set members closest to the test sample; but classification based on the weighted vote of these K-nearest neighbors, seems closer in spirit to non-parametric probability estimation. Both these classification methods are explored more fully in Appendix B.

III. PRACTICAL CONSIDERATIONS IN THE IDENTIFICATION OF NON-COOPERATIVE SIGNALS

The foregoing discussion of the basic techniques of mathematical pattern recognition is intended to serve as an outline of approaches to the general problems of classification. Identification of the originator of a banded signal, while bearing considerable similarity to Bayesian classification, is complicated by a number of factors not normally arising in such problems as the identification of handwritten characters.

In many of the classical problems of mathematical pattern recognition, samples are obtained from a statistical universe which is well-defined in the sense that the number of classes to which the samples may belong is finite and previously ascertained. Additionally, the parameters or features used to typify these classes are normally time-invariant. The set used for training is accurately classified and sufficiently large that meaningful statistical estimates of the parameters and their distributions may be obtained.

Under these conditions, the processes of Section I find their greatest application. Unfortunately, they are seldom encountered in identifying the signals of a non-cooperative originator. The data base used for a training set does not normally provide arbitrarily accurate ground truth classification. In practice, even the number of

actual originator classes may not be known. In many cases, the paucity of signal intercept precludes more than the grossest statistical parameter estimation. The parameters themselves are time-dependent since they are affected by component aging, time-dependent signal propagation conditions, and unknown external maintenance and adjustment. The advantages of long-term averaging to obtain statistical convergence, i.e. accurate a-priori estimates, may be obscured by short-term perturbations for which adequate correction may be prohibitively costly or undefinable.

In summary, the investigator proposing to use pattern recognition techniques to identify the originator of hostile signals is faced with a series of constraints involving small sample size, time-varying signal parameters and class probabilities, inaccurate ground truth, and an indeterminate number of classes with which signals may be associated. Such constraints impose severe limitations upon the pattern recognition techniques which may be successfully applied to the problem. The following sections indicate some of the procedures which deal with these limitations.

A. SMALL SAMPLE SIZE

The difficulties arising from small sample size are most immediately apparent when one attempts to increase the number of features in order to facilitate the separation of classes. The dominant effect is one of inaccuracy arising from under-sampling the required d-dimensional a-priori probability

distributions. When this undersampling occurs, the most likely result is that increasing the number of features becomes counterproductive to "good" classification. In the practical multi-category problem, it is not unusual to encounter an initial set as many as fifty features which the designer believes to be useful in identification. The number of training samples is often too small to allow meaningful Parzen window or other non-parametric density estimation techniques over this number of dimensions.

An alternative procedure of characterizing the a-priori distribution function by use of the most general multivariate normal form requires estimation of a " $d \times d$ " non-singular covariance matrix. This imposes an algebraic requirement of at least $d+1$ samples. However, experience in the practical estimation of covariance matrices suggests that at least two to three times this minimum algebraic requirement of $d+1$ samples is preferable [7, 11]. Frequently the number of available training samples for such complete estimation of the individual class covariance matrices is inadequate.

Two common methods of correcting this difficulty involve somewhat arbitrary assumptions about the nature of the class covariance matrices. One possible assumption is that all classes share the same covariance matrix \underline{S} , which is then obtained by averaging the individual class covariance matrices \underline{S}_i :

$$\tilde{S} = \frac{1}{\sum_{i=1}^m K_i} \sum_{i=1}^m \frac{K_i}{(K_i - 1)} \sum_{j=1}^{K_i} (\underline{x}_j^{(i)} - \underline{\mu}^{(i)}) (\underline{x}_j^{(i)} - \underline{\mu}^{(i)})^T \quad (\text{III-1})$$

A second method assumes statistical independence of the features and makes all off-diagonal elements zero, regardless of evidence to the contrary. This approach preserves additional detail about individual class structures at the expense of disregarding highly correlated features. It requires only the calculation of a single feature class means and variances, as indicated in Equations A-1 and A-2. Although such assumptions are almost surely incorrect, they often lead to better classifier performance than a maximum likelihood estimator [7].

Small sample size also leads to complications when one attempts to estimate the error rate of the classification algorithm finally adopted. If one chooses to estimate the classifier's error rate from the assumed parametric model, one finds the result is optimistic to the extent that the training samples are peculiar and unrepresentative [21].

An empirical approach to determining error rate avoids this problem by running the classifier on a set of test samples. Where one is faced with a small number of available signals for which ground truth is accurate, the confidence which may be placed in the results of the empirical scheme is marginal. For example, if two errors are made in ten

test samples, one may predict with 95% confidence that the true error rate of the classifier lies somewhere between five and fifty-five percent. The figure of ten test samples is not unrealistic if one uses different design and test samples to avoid the hazards of "testing on the training data."

B. INDETERMINANCE OF CLASSES

The absence of a predetermined number of classes presents additional problems which must be addressed in any practical identification scheme. The first problem is that of establishing the criteria for excluding a given sample as a member of any of the previously determined classes. The second problem is one of associating or "clustering" the resultant "unclassified" signals into possible new classes.

Within the formal structure of minimal risk Bayesian classification, one may introduce the concept of an additional class which corresponds to the heuristic category of "don't know." More frequently, this category is interpreted as that portion of the feature space for which no feature-conditional class probability exceeds a given threshold. That is:

$$p(C_i | \underline{x}) < \lambda \quad \text{for all } i \quad (\text{III-2})$$

An alternative method of establishing this threshold for multivariate normal distributions is to specify a

maximum Mahalanobis distance from the established class means.

$$r_i = (\underline{x} - \underline{\mu}^{(i)})^T \underline{S}_i^{-1} (\underline{x} - \underline{\mu}^{(i)}) \geq \lambda \quad (\text{III-3})$$

as the boundary for which a sample will be included in a given class.

The second problem, that of clustering the "unclassified" signals, is less well defined from the standpoint of Bayesian analysis. It represents one of a group of problems known as unsupervised learning. Jarvis and Patrick [12] present several techniques by which such clustering may be performed and illustrate the advantages in graphically displaying the clustering process.

One such clustering technique used in several identification systems is illustrated for a two-dimensional case in Figure 2. Thresholds based on Mahalanobis distance are established for each cluster. If the Mahalanobis distance for a given sample, such as S_1 , exceeds the outer threshold distance from the mean of any known cluster, that sample is considered to represent a new cluster and possibly a new signal originator. For this new cluster, arbitrary distance thresholds based on "typical" class covariance are assumed. Subsequent samples (S_2) lying within the inner threshold are used to update the class statistics; those samples (S_3) lying between the two thresholds are tentatively associated with the class, but do not update the class statistics.

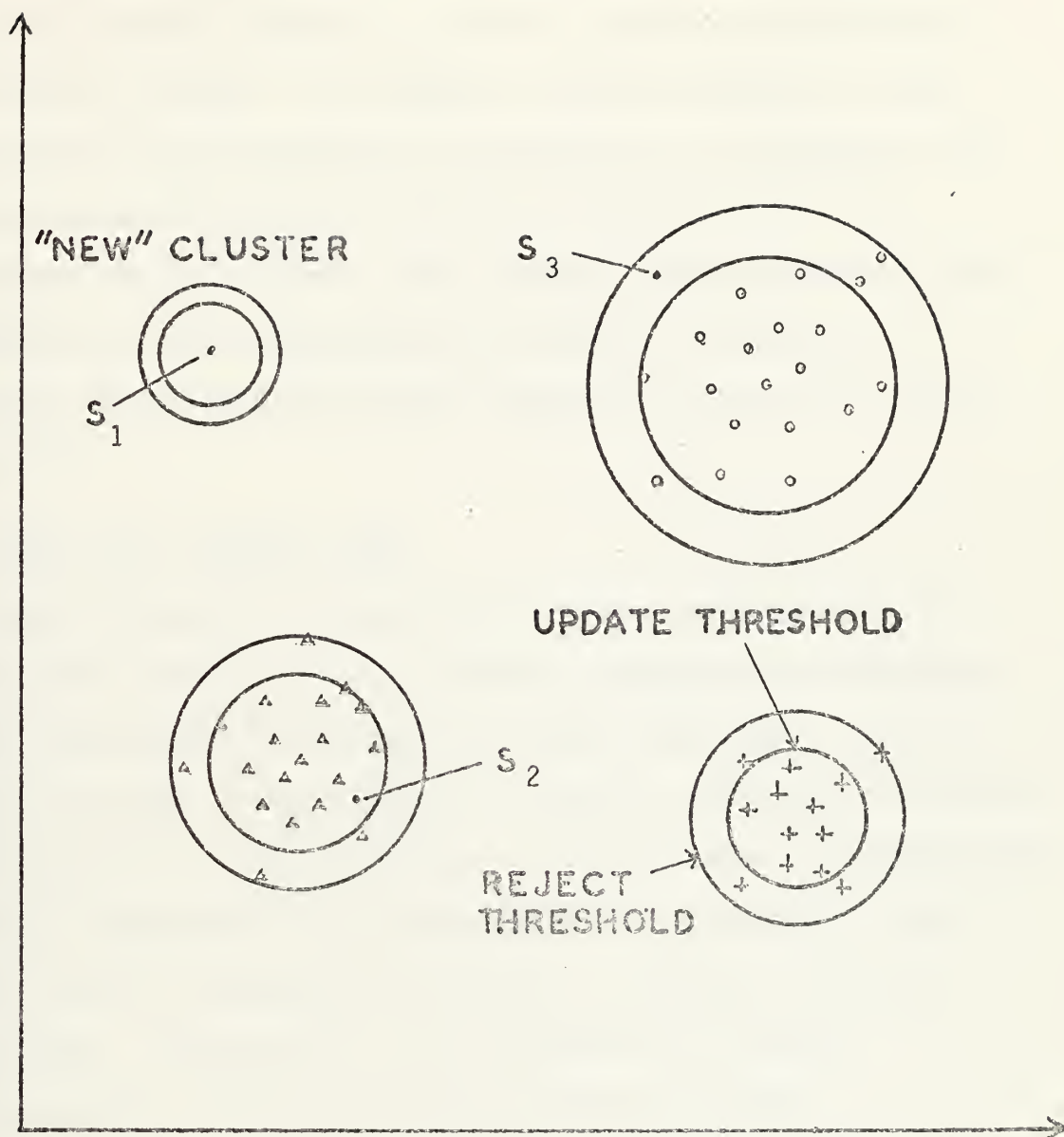


FIGURE 2. Guard Zone Clustering

It is during the clustering procedure that the man-machine interface becomes a critical factor in the pattern recognition process. An equally important factor in the efficiency of the clustering algorithms is a technique for feature selection which provides an adequate basis for representing new classes. The initial investigator has the obligation to use a training set which as nearly as possible represents the entire spectrum of possible feature parameters.

C. INACCURATE GROUND TRUTH

The difficulty of obtaining accurate correspondence between the elements of the initial training set and known signal originators represents the most persistent and perplexing dilemma in the design of pattern recognition devices. Those signals for which there is a high degree of probability that the originator can be determined from external information may not adequately represent the entire spectrum of signals. Conversely, the originators of a more representative group of signals may not be known.

One may attempt to choose self consistent classes through the use of a clustering algorithm and the application of probability estimator to information confirming class membership, a technique best described as "learning with a probabilistic teacher." This approach, discussed more fully by Cooper [4], is still rather exploratory in nature and is implemented at considerable cost in timeliness and computational efficiency.

Since the empirical methods by which any new classification performance is evaluated must use as ground truth the identifications of an older system, indications of any improved performance in the new system are perforce intuitive. This is the crux of the circular dilemma which arises when one is forced to estimate not probability of error, but probability of disagreement. In the event of disagreement, which system is right?

IV. FEATURE SELECTION IN WAVEFORM ANALYSIS

One important question regarding the process of waveform identification has not yet been addressed. All of the techniques described so far have indicated that one commences this process with a particular set of feature measurements. In the case of a waveform $z(t)$, a sufficiently general feature vector might be the sequential samples of the continuous signal, i.e. $\{z(t_1), z(t_2), \dots, z(t_n)\}$. If these samples occur at greater than the Nyquist rate, one may reasonably assume that they constitute an adequate description of the signal. One could then proceed directly to the calculation of n -dimensional probability distributions for each class, blithely oblivious to the fact that n may be on the order of 500 or more.

In light of the advantages in the choice of a small number of mutually-independent well-separated features for the process of pattern identification, it is normally advisable to seek some mapping, F , of the signal space Z into a pattern space X . That is, we wish to find a preprocessing transformation so that the pattern space $X = F(Z)$ satisfies the following objectives:

- (1) low dimensionality
- (2) retention of sufficient information
- (3) enhancement of distance in pattern space as a measure of the similarity of physical patterns

(4) comparability of features among samples.

Note that 1 and 2 imply elimination of redundant information.

In addition to the above objectives one might also wish to obtain a pattern space representation adequate for the construction of new classes in an unsupervised learning mode. To this end, it is convenient that the individual components of the pattern space have some natural interpretation which might provide qualitative information about the underlying causes of class difference.

As a first step toward eliminating redundant information in a pattern space satisfying the above criteria, it is frequently desirable to represent the sampled waveform as a linear combination of orthonormal functions. Two commonly used orthonormal expansions satisfying slightly different training set mean square error criteria are discussed in Appendix C. The coefficients of these functions may then be used as features of the signal. The choice of orthonormal functions used in the expansion however, need not be one which necessarily minimizes a form of mean square error in the representation of the signal. If one is willing to accept a slightly greater amount of error in modeling the training set waveforms, then one possible technique is to use a set of orthonormal functions for which it is computationally efficient, as it is with the fast Fourier transform, to obtain coefficients. By truncating the number of coefficients one may achieve some reduction of dimensionality.

The features derived from the orthonormal projections of the sampled signal described in Appendix C represent linear transformations of the sampled waveform data. For the purposes of signal identification, this representation may not be adequate.

Since one of the goals in this application of pattern recognition techniques is to assist a human analyst in making an identification, it is particularly desirable that the resultant features derived from preprocessing have some degree of natural interpretation. The optimal linear transformations discussed do not always lend themselves to easy extension in an unsupervised learning mode.

Features which are more directly related to an assumed model of the signal generating process and which are often easily implemented as a measurement procedure, may in fact involve non-linear transformations of the waveform data. In selecting features of this type, the prior experience of the data analyst and the details of the signal generating model serve as a focus for the development of non-linear preprocessing.

Quite commonly, the initial feature space of the signal will include components derived from readily interpretable data transformations, both linear and non-linear in nature. This feature set is normally too large for convenient computation and some of the features thus obtained may be dependent. The techniques of linear dimensionality reduction

described in Appendix C may be applied to this set of features, particularly when graphical display of the individual signal is required. In order to provide a feature set with dimensionality low enough to insure computationally convenient classification and general enough to allow the establishment of new class clusters, it is both sufficient and desirable to reduce dimensionality while retaining the identity of individual features by selecting those n of m features which lead to the optimal separation of classes. The selection process then requires both the adoption of a criterion by which class-separability may be estimated and the development of a feature search algorithm. Both of these topics are discussed in Appendix C.

V. PARAMETER MAPPING OF RASTER SCAN DISPLAYS

As a practical application of the signal classification techniques previously described, the author conducted a series of investigations which used as a data base the raster scan displays of bauded, frequency-shift-keyed (FSK) signals transmitted by several originators. The signal data base, many of the associated processing algorithms and the processing equipment itself were made available through the cooperation of Electromagnetic Systems Laboratories in Sunnyvale California. The data base and supporting facilities are the results of the company's development of the Parameter Encoder in a Navy-sponsored program for signal measurement and identification.

The objective of the author's investigation was to isolate from the raster scan data a small set of features which might prove useful in automated signal identification. As a part of the existing clustering and identification technique, the raster scan pattern of a given signal was presented to the system operator for visual comparison with patterns of previously identified signals. It was claimed that such comparison was of value in the final stages of the identification process, when the classification of the signal in question had been narrowed to a few possibilities. Display of the information necessary to produce the required number of raster scan patterns, however, consumes a significant

amount of computer time. The interests of improved efficiency, coupled with the author's desire to quantify a process open to considerable latitude in operator judgment, provided the motivation for the project.

In obtaining a reasonable initial set of parameters for consideration by this thesis, a study of the signal generating process and the characteristics of the raster scan display led to a slight modification of the normal minimal phase raster representation. The resulting "phase-unwrapped" raster display was the basis of a series of measurements related to clockrate, transient phenomena, and possible data-dependent or unintentional external rate modulation. A model of the signal process, the raster scan display, and the methods used to measure the initial set of features are discussed in the following sections.

A. SIGNAL GENERATION PROCESS

Since the raster scan display is based essentially on zero crossing information, it represents a non-linear transformation of the incident signal. It is convenient to refer to a model of the signal generating process in an attempt to justify use of such a transformation. From this model the qualitative effects of the various steps in the generation of the raster scan pattern can be estimated. Fig. (3) provides an outline of the process which leads to the raster scan display. All of the signal data base information available for the feature measurements proposed by this thesis is

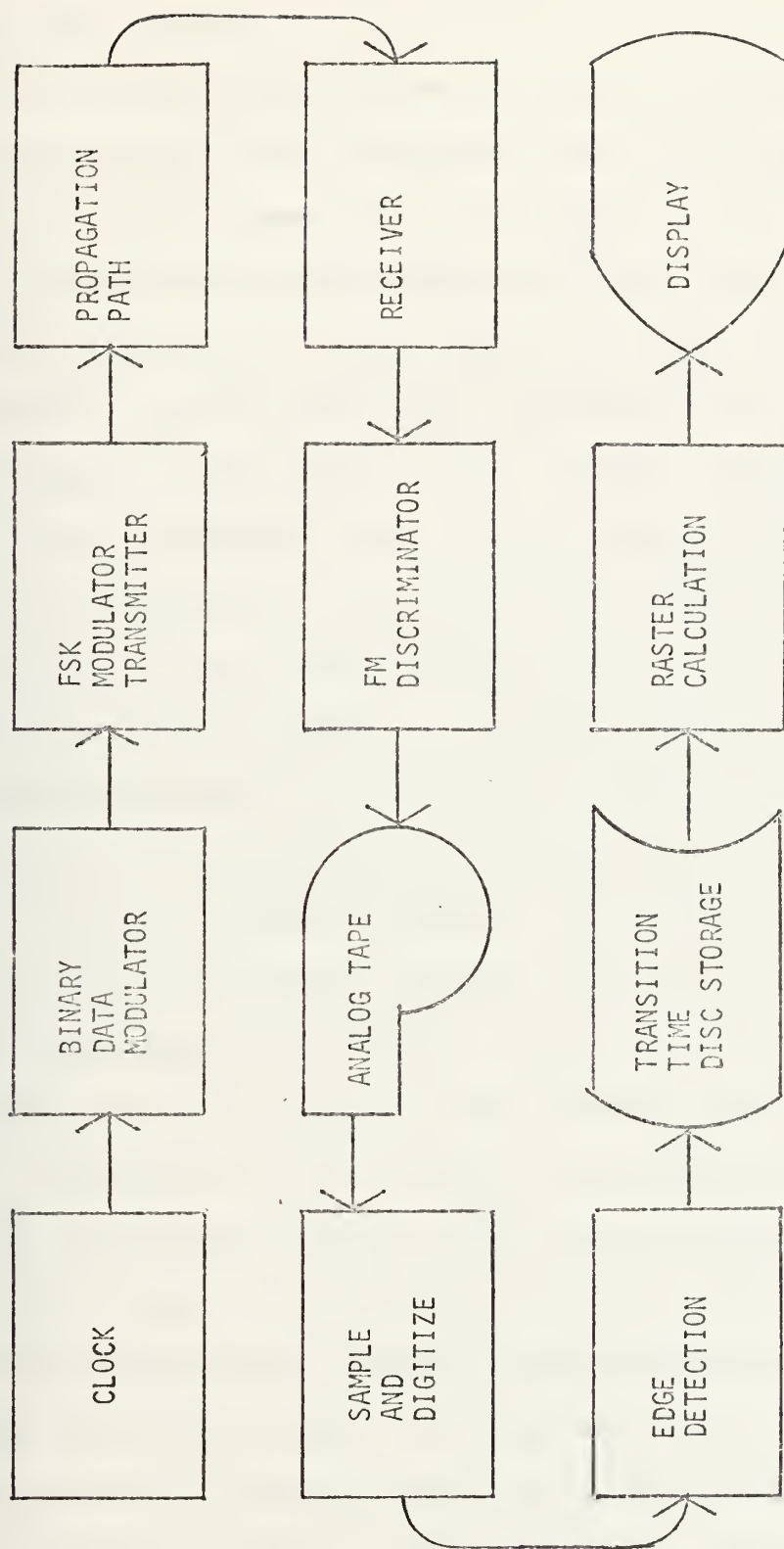


FIGURE 3. Raster Signal Generation Process

contained in the transition time measurements provided by the edge detector.

Of all the qualitative effects upon transition time measurements, those associated with the originator's clock and bit-stream generator are potentially of greatest value to a raster-scan-based identification scheme. Class-to-class differences in the basic clock rate are certainly of interest, as are clock rate variations due to component instability and unintentional external rate modulation. If the originator's clock is not continuously running, transient phenomena associated with clock turn-on may be apparent in the raster display. The binary bit stream generator (data modulator) and non-linearities in the FSK modulator itself may produce an unequal mark-space duty cycle or data dependent mark-space asymmetries which appear as bias in the raster display. Such effects should produce features of the raster display which serve to characterize the originator. To isolate them effectively, however, may require the use of non-linear waveform data transformations.

Other parts of the signal generating process may introduce qualitative effects which tend to obscure those produced by the source. For example, noise and signal attenuation in the propagation path may introduce FM spikes and anomalous transition times. Non-linearities and zero-level offset in the FM discriminator may produce additional bias in the measured mark and space durations. The waveform sampling

procedure itself introduces an unavoidable time quantization error which may or may not be uniformly distributed on a short term basis. If this error is uniformly distributed, one can expect a transition time error variance σ_T where

$$\sigma_T = \frac{\Delta t}{12} \quad \text{V-1}$$

and where Δt is the sampling period.

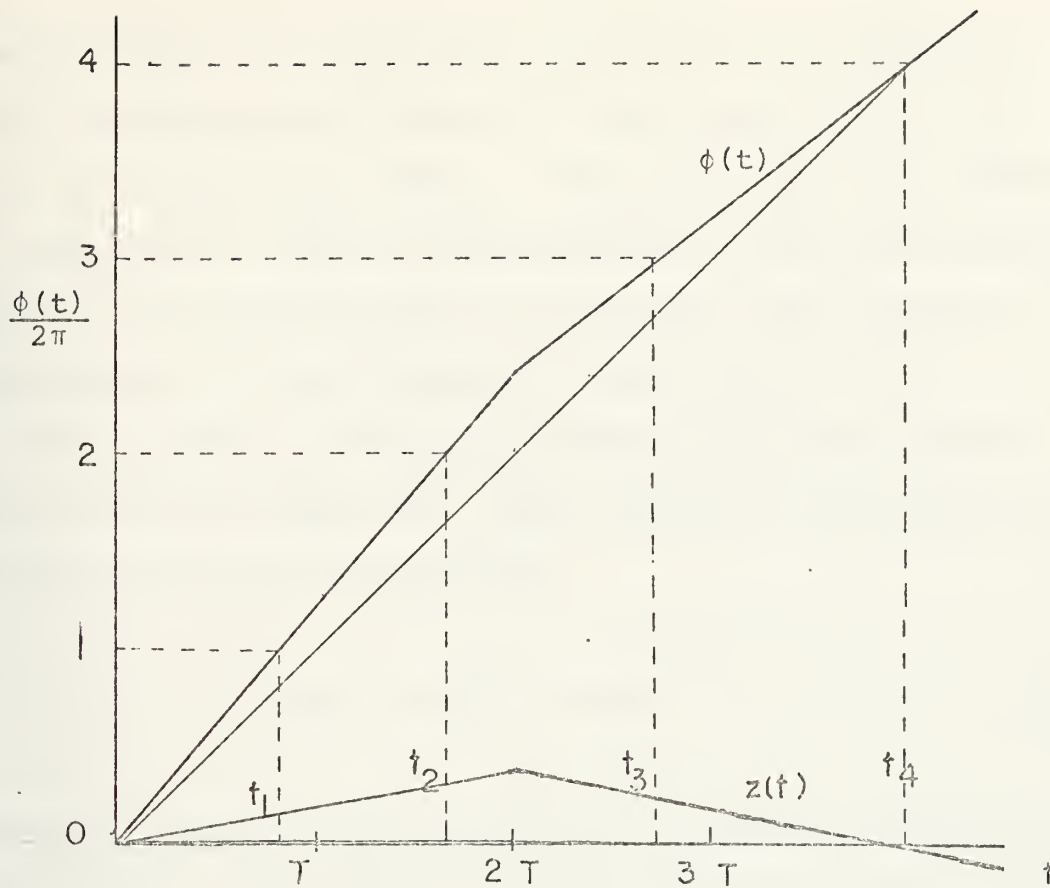
B. RASTER SCAN DISPLAY

The use of a raster scan display arises rather naturally in the attempt to represent the fine details of clock rate variation. The axes of the display correspond to coarse and fine divisions of time in the following manner. Suppose some event, such as a mark-space transition, is assumed to occur only at times characterized by some integer multiple of a nominal period T . The actual time of occurrence, t , may then be represented by

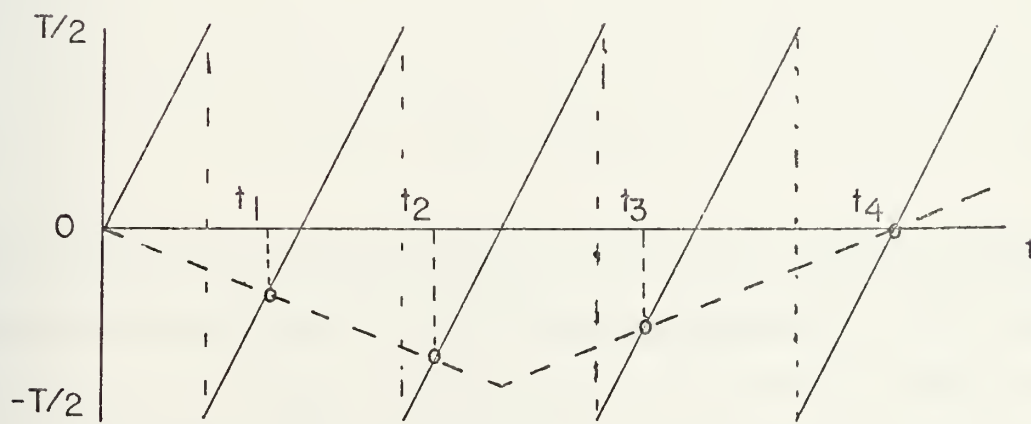
$$t = nT + z$$

$$-T/2 < z \leq T/2; \quad n \text{ integer} \quad \text{V-2}$$

By equating the horizontal axis of the raster display to the actual time variable, t , and the vertical axis to the fine time z , any arbitrary time may be represented as a point on the sawtooth waveform indicated in Figure 4b. The



(a) Phase and Phase Perturbation vs. t



(b) Raster of Events t_1-t_4

FIGURE 4. Relation of Phase Perturbation to Raster Display

raster display has the effect of mapping the points t_i in a one-dimensional region into the points $(z(t_i), t_i)$ of a two-dimensional region. Since the fine time variable, z , represents a lag or lead relative to the nominal period, one can relate the points of the raster scan display to the samples of some continuous phase perturbation.

For example, consider the continuous phase function illustrated in Figure 4a. The function is comprised of a linear part and a perturbation

$$\phi(t) = 2\pi\left(\frac{1}{T}t + z(t)\right) \quad V-3$$

where T is the nominal period of the raster scan display at times t_i . If samples of this function are taken when

$$\phi(t_n) = 2n\pi \quad n = 0, 1, 2, \dots \quad V-4$$

and if

$$|z(t)| < T/2 \quad \text{for all } T \quad V-5$$

then the minimal time difference from the expected sample time is proportional to the phase perturbation. The raster scan points of Figure 4b represent non-uniform samples of this continuous perturbation. In this sense one can represent the events occurring at times (t_1, t_2, \dots) as samples

($z(t_1)$ $z(t_2)$...) of a continuous phase proportional function $z(t)$ where

$$z(t_n) = t_n - n_{\max} T$$

where

$$n_{\max} = \left[\frac{t_n}{T} + \frac{1}{2} \right] \quad \text{V-6}$$

i.e. the greatest integer less than $\left[\frac{t_i}{T} + \frac{1}{2} \right]$.

Figures 5a through 5d illustrate the raster patterns of events t_n of arbitrary starting phase and periodicity different from the nominal period of the raster scan. If the nominal raster period coincides with the actual periodicity, as in Figures 5a and 5b, the slope of the line joining points of the raster pattern is zero. Similarly, if the periodicity of the series of events differs slightly from the nominal raster period by a small amount ϵ , the raster pattern may be characterized by a line whose slope is given by:

$$\frac{z_2 - z_1}{t_2 - t_1} = \frac{\epsilon}{T + \epsilon} \quad \text{V-7}$$

as shown in Figures 5c and 5d.

In the signal display of the Parameter Encoder both mark-space and space-mark transition times are plotted on a

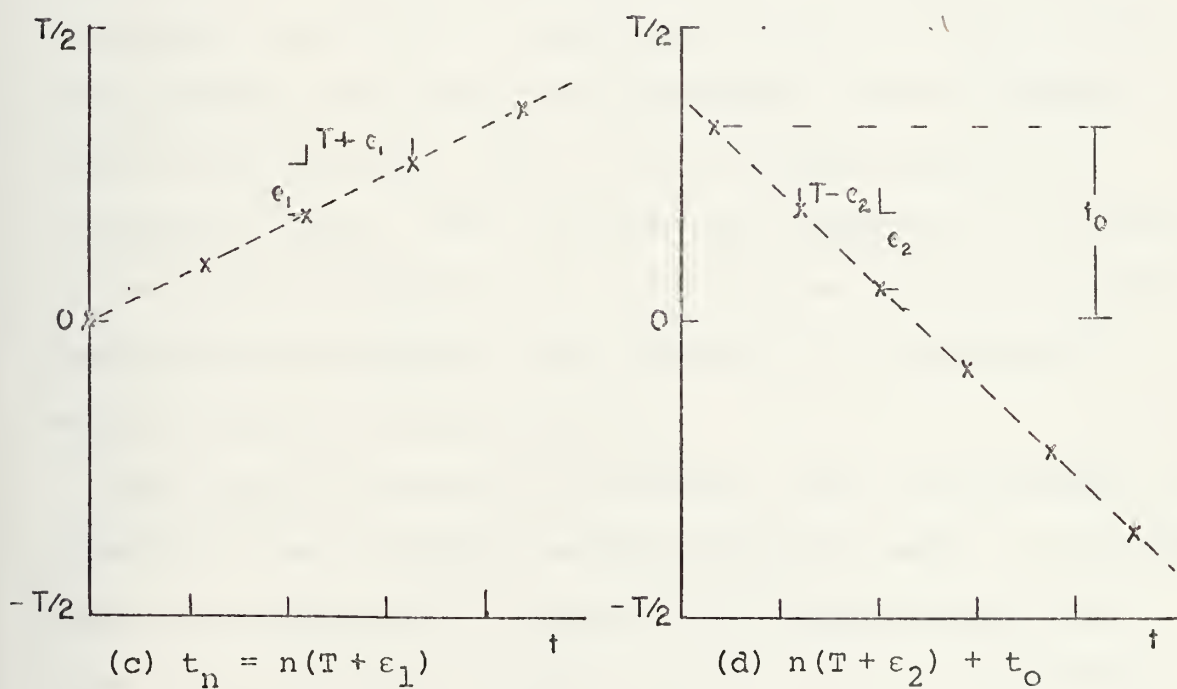
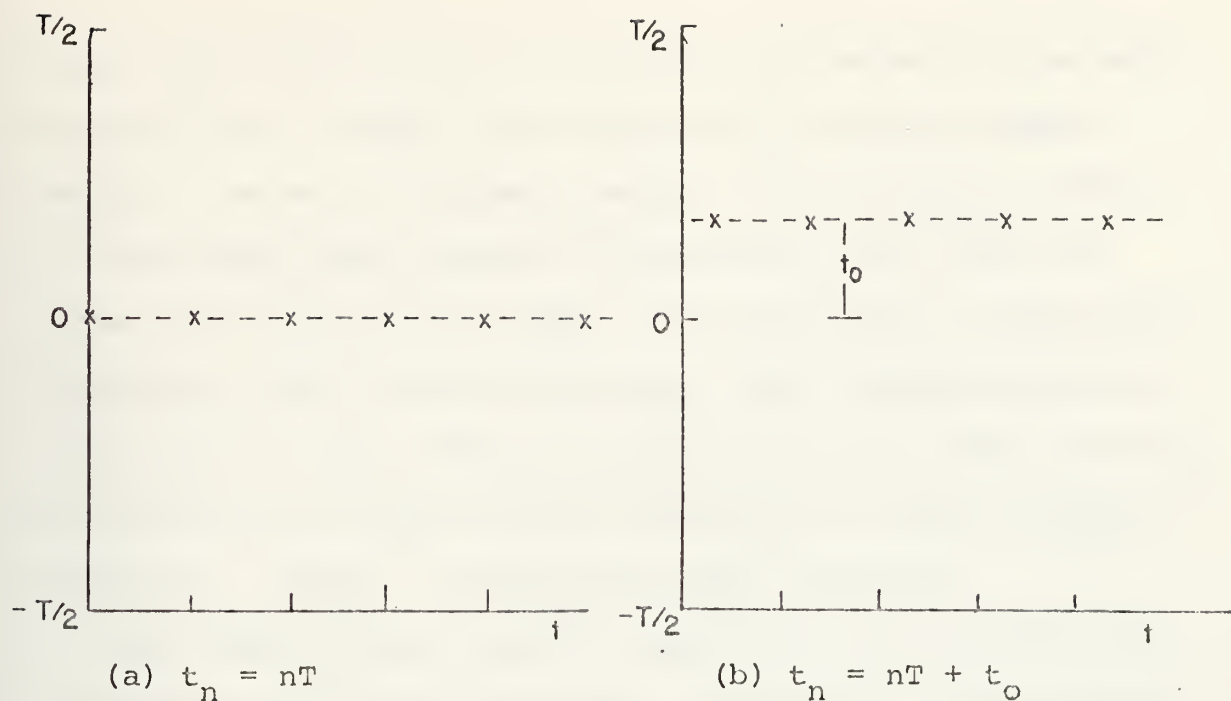
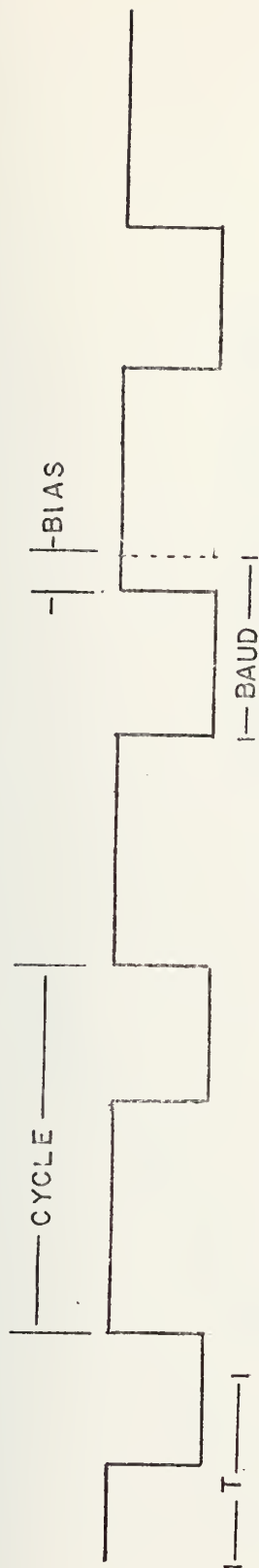


FIGURE 5. Raster Display of Constant Period Events t_n

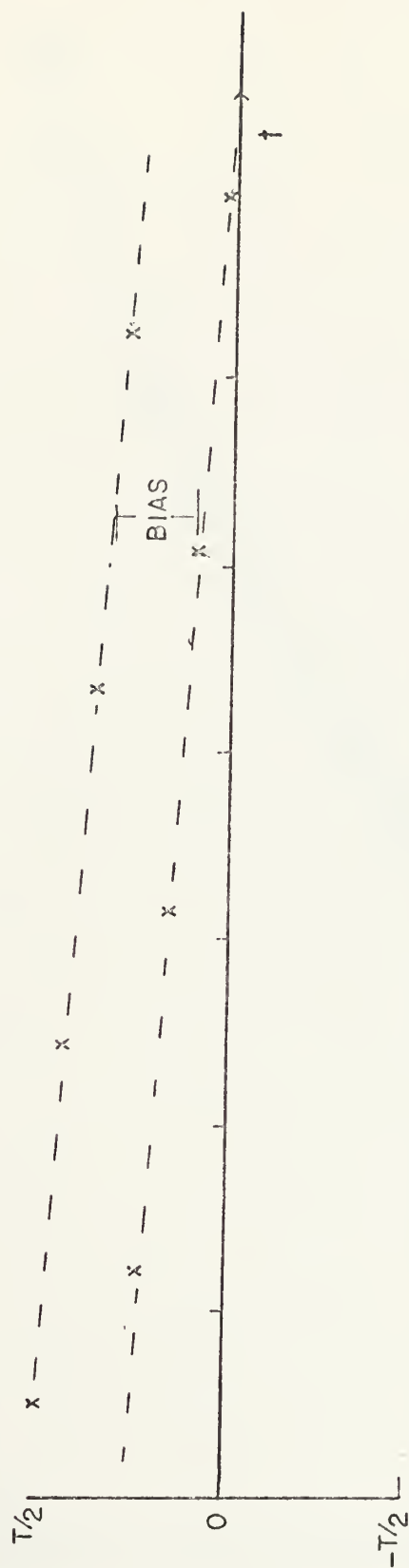
raster scan whose nominal period is approximately one-half the time for a single mark-space cycle, that is raster period is equal to the baud length. The alternate points of the raster scan exhibit a bias term if the mark and space durations are not equal. If the period of the entire mark-space cycle remains constant, lines connecting alternate points of the raster scan have the same slope, but they are vertically offset by the difference in mark and space durations. Figure 6 illustrates this condition.

The effect of the uniform sampling of the waveform and the subsequent time quantization of events is apparent when the nominal period of the raster scan is on the order of ten times that of the sampling period. Figures 7 and 8 illustrate the effect of quantization noise in two raster scan patterns of a slowly varying signal, one of which uses a nominal period that is an integer multiple of the sampling time. This quantization phenomena was evident in the raster displays of the signal data base used in this thesis since the nominal baud length is on the order of twenty times the sampling period.

The abrupt transition in Figures 7 and 8 also serves to illustrate the "aliasing" effect which may occur in a minimal phase representation if the time difference between the actual and the nominal event exceeds one-half the nominal period. When this aliasing occurs the points plotted on the raster scan are no longer proportional to the assumed



(a) Biased Signal Waveform



(b) Raster of Biased Signal

FIGURE 6. Raster Display of Bauded Signal Bias



FIGURE 7. Quantization and Aliasing Effects in Raster Display
 Nominal Period = $18.6 \times$ Quantization Interval

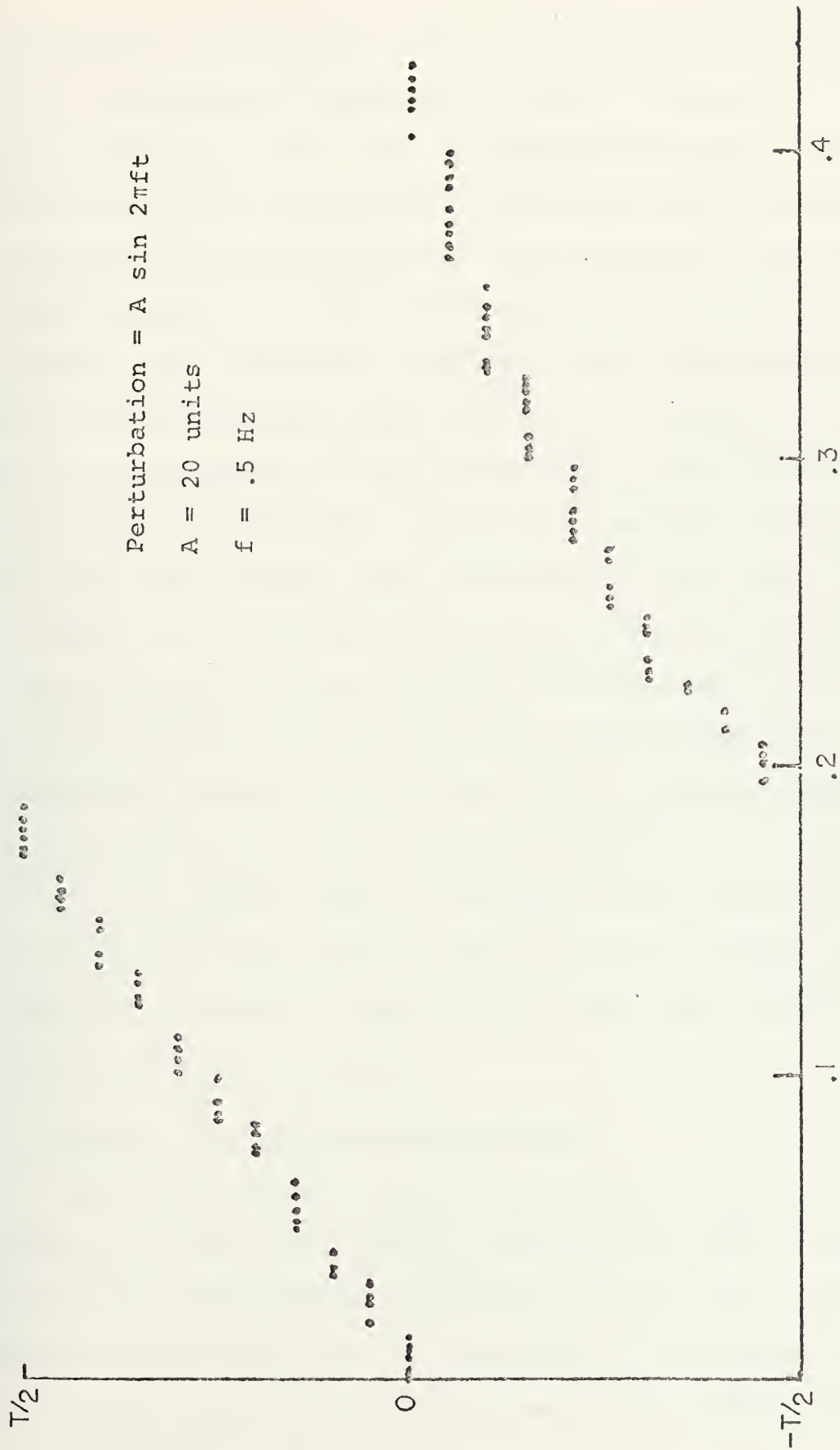


FIGURE 8. Quantization and Aliasing Effects in Raster Display
Nominal Period = 20 Quantization Units

continuous perturbation. In order to maintain proportionality it is preferable to represent the time of events by a minimal phase difference plot. Such a procedure performs a certain amount of "phase unwrapping" by assuming that the absolute raster time difference between two successive points never should exceed $T/2$. Figure 9 illustrates the effect of this minimal phase difference algorithm. Since each segment of the sawtooth pattern raster scan may be extended in either forward or backward in time the minimal phase difference representation uses this extension to represent those points for which the minimal phase difference is less than $T/2$. In Figure 9, those points occurring in the heavily lined region indicate the normal raster scan pattern while a continuous line joins the points of the minimal phase difference pattern. Since some of the "aliasing" associated difficulties in the measurement of features may be eliminated by use of a minimal phase difference, this representation was included as the initial step in the measurement process. Figure 26 outlines the flow of the algorithm actually used in the measurement program.

C. RASTER SCAN DESCRIPTIVE PARAMETERS

The first efforts of the author's investigation of the raster scan data were directed towards obtaining a rather large set of descriptive parameters, particularly those which seemed useful in the quantitative representation of the effects anticipated from a study of the model signal

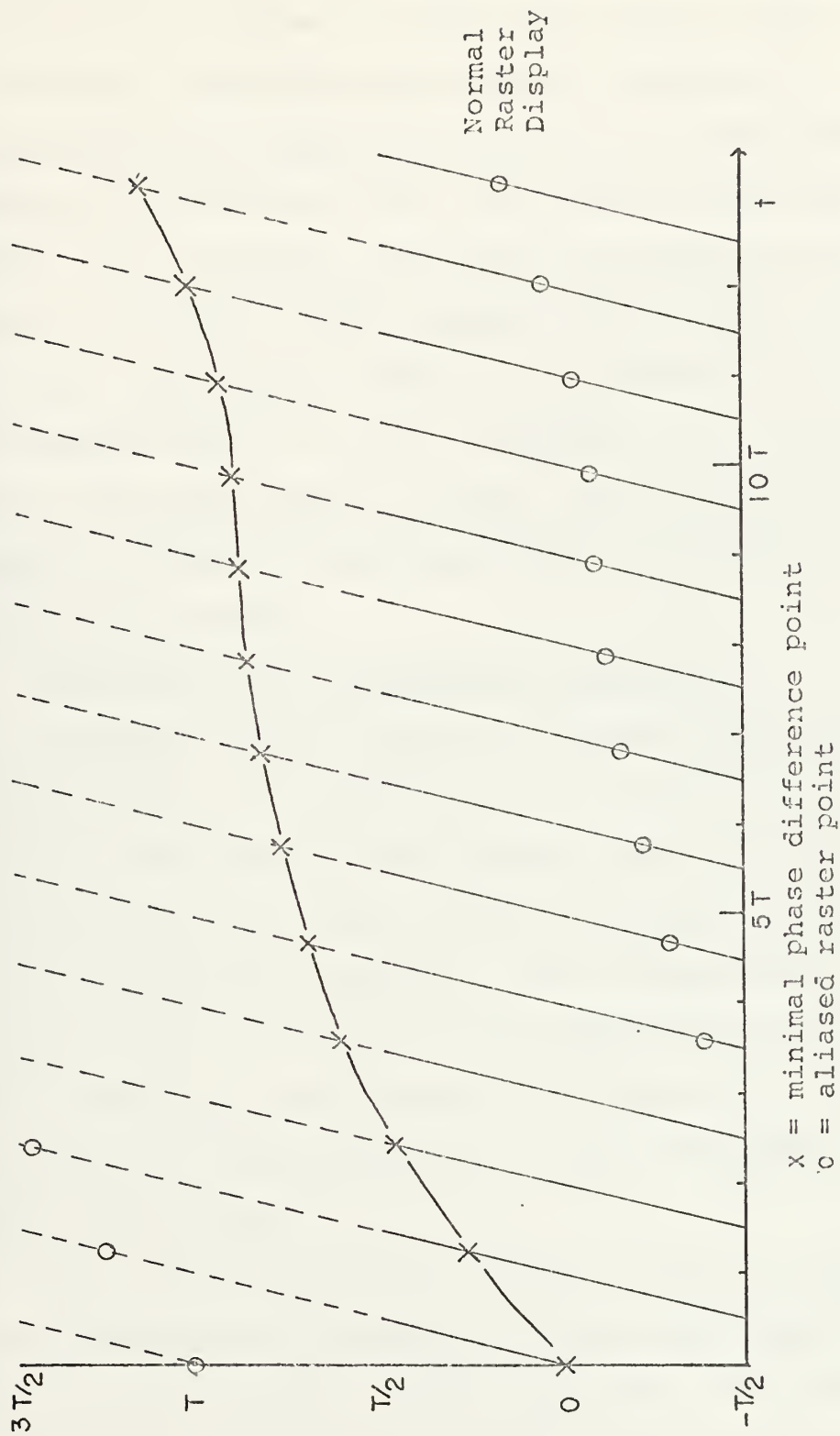


FIGURE 9 Minimal Phase Difference Raster Display

generating process. Manipulations of the transition time data to provide information about the nominal clock rate, it's transient variations and external modulations, as well as source-associated mark-space bias and data-dependent transition time anomalies were the areas of particular interest. The parameters considered included representations of all or part of the signal's raster phase by polynomial mean square fits, Laguerre polynomials, and Fourier power spectrum terms, as well as measures of mean bias, intrinsic and total signal variance. The techniques employed to calculate these parameters are described in the following sections.

1. Mean Bias and the Effect of Transients

Although several sources of apparent mark-space bias other than those associated with the originator have been indicated in the study of the process model, the measurement of this parameter may provide an adequate estimate of the source-associated component. More recent studies of new signal data for which the bias effects not associated with the originator are believed to be negligible, indicate that such source-dependent bias not only exists but can in fact be time-varying.

The simplest estimate of mean bias may be obtained by the straightforward averaging of the raster phase difference of alternate points on the raster scan. For the raster scan display of the alternate positive and negative transitions,

$z^-(t_1), z^+(t_2), z^-(t_3) \dots, z^+(t_{2n})$, the mean bias \hat{b} is given by:

$$\begin{aligned}\hat{b} &= \frac{1}{n} \sum_{i=1}^n z^+(t_{2i}) - z^-(t_{2i-1}) \\ &= \frac{1}{n} \sum_{i=1}^n z^+(t_{2i}) - \frac{1}{n} \sum_{i=1}^n z^-(t_{2i-1}) \\ &= \hat{z}^+ - \hat{z}^- \qquad \qquad \qquad \text{V-8}\end{aligned}$$

which is the difference in the mean raster phase of the two bias conditions.

One difficulty which arises from this representation of mean bias occurs as a result of clock transient behavior. If at any point t_i , raster phase $z(t_i)$ differs from the raster phase of the starting point by an amount approximately equal to one-half the nominal raster period, the cyclic representation (minimal phase) of the raster display introduces an "aliasing" error into the mean bias estimate of Equation V-8.

To eliminate such errors in the bias estimate, one must use the minimal phase difference representation discussed previously for each bias condition.

If the transient behavior of the clock rate is sufficiently violent, even the minimal phase representation may not be adequate. For example, assume that the instantaneous

originator clock frequency f , is characterized by turn-on from a dead stop:

$$f = f_{ss} [1 - \exp(-at)] \quad V-9$$

where a is the reciprocal of turn-on constant, τ . The instantaneous phase $\phi(t)$ in cycles is obtained by integrating f from 0 to t :

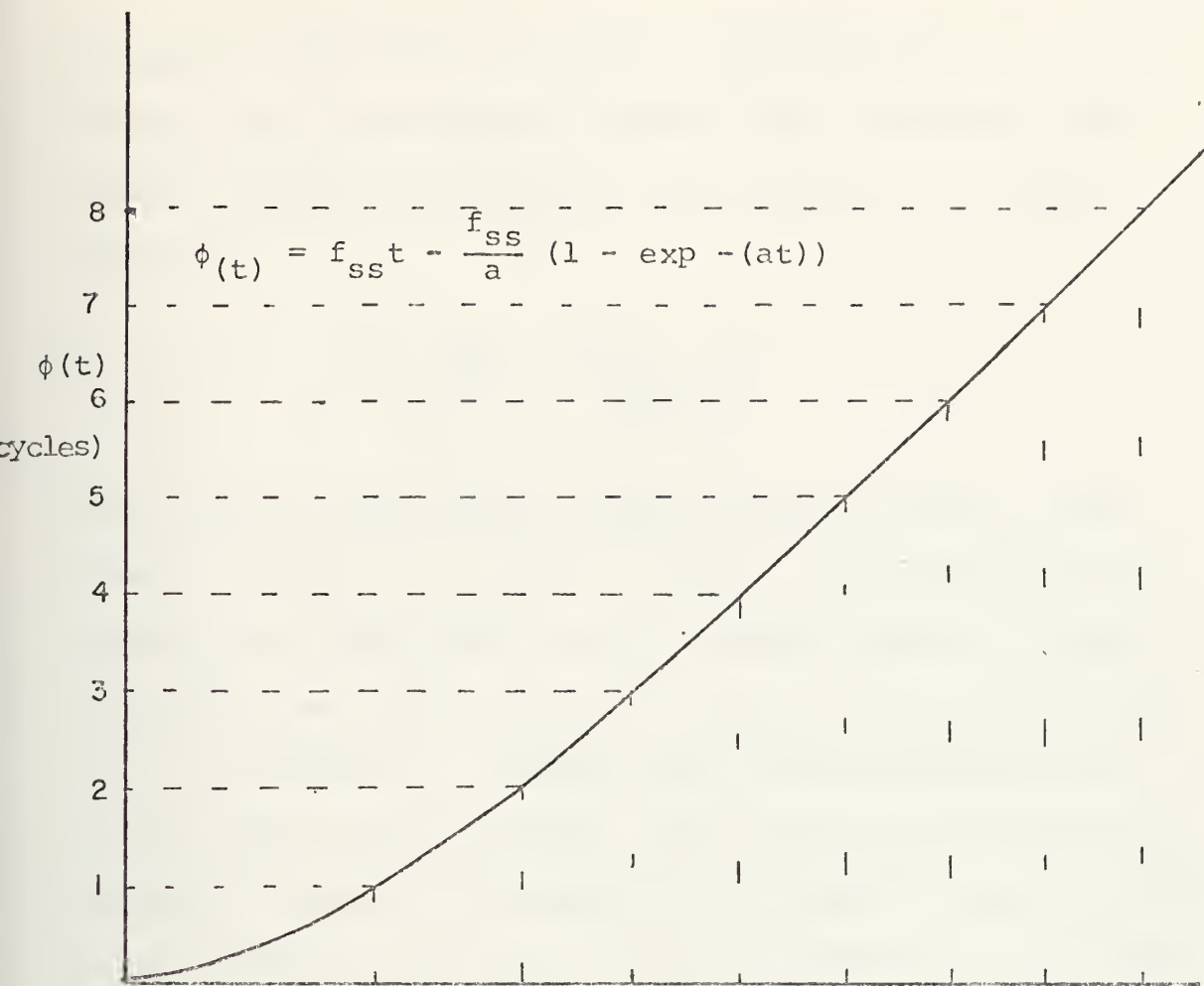
$$\phi(t) = \{f_{ss} t + \frac{f_{ss}}{a} [-1 + \exp(-at)]\} \quad V-10$$

If transitions occur when $\phi(t) = 0, 1, 2, \dots$, (Figure 10a), the raster scan points of Figure 10b are obtained. It should be noted that the transition times are non-uniform samples of the actual phase. Similar computations may be performed for the raster phase of a clock whose transient frequency behavior is characterized by a shift from a free running frequency, f_o , to the steady state frequency, f_{ss} .

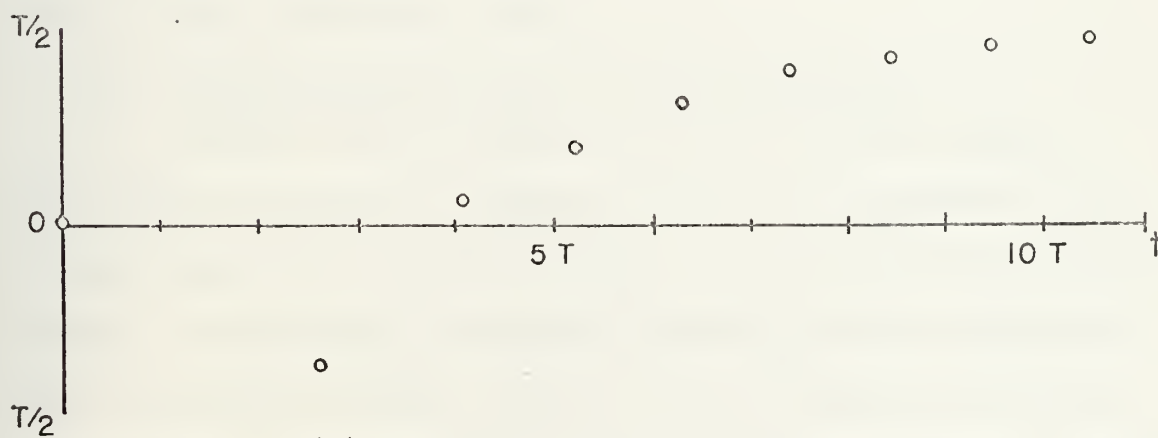
$$\begin{aligned} f_o &= f_o \exp(-at) + f_{ss} [1 - \exp(-at)] \\ &= [f_o - f_{ss}] \exp(-at) + f_{ss} \end{aligned} \quad V-11$$

$$\phi(t) = \frac{[f_o - f_{ss}]}{a} [\exp(-at) - 1] + f_{ss} t \quad V-12$$

The above results suggest that the raster-phase pattern or a minimal phase difference pattern may be



(a) Phase vs. Time



(b) Resultant Raster Pattern

FIGURE 10. Relation of Exponential Transient to Raster Display

extremely misleading during the transient period. It is evident that the predicted raster phase (in cycles) may differ from the actual steady state phase by an amount equal to

$$\frac{f_o - f_{ss}}{a} = \frac{T_{ss} - T_o}{T_o T_{ss}} \tau$$

which may be considerably greater than one cycle. For example, if $T_o = .9t_{ss}$ and $\tau = 20 T_{ss}$, then difference in steady state phase may be more than two complete cycles. Using the same time constant τ , turn-on from a dead stop ($f_o = 0$) results in a steady state phase difference of twenty cycles (from nominal). The raster scan display of Figure 11 serves to illustrate the transient turn-on effects observed when $\tau = 10 T_{ss}$. In such a case the anti-aliasing capability of the phase unwrapping process is severely strained and it would be wise to avoid measurement of mean bias in this transient region.

2. Pattern Variance and Intrinsic Variance

Pattern variance and intrinsic variance measure the extent to which the raster pattern is explained by the nominal clock rate and the minimal phase difference representation. Large values of total pattern variance indicate that the nominal raster chosen does not completely explain the pattern. Large values of the intrinsic variance suggest that a minimal phase difference representation is not

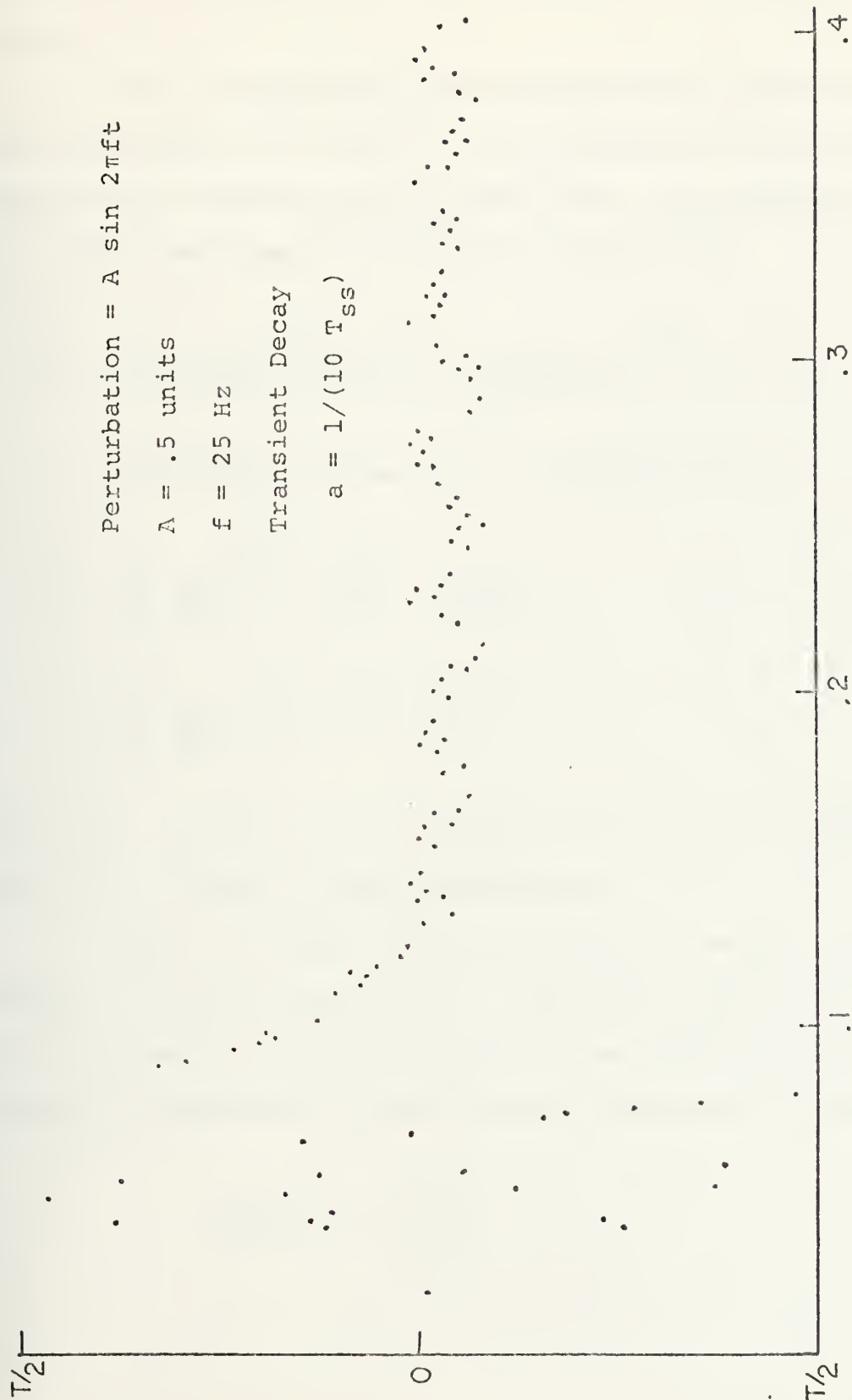


FIGURE 11. Effect of Transient on Normal Raster Display

adequate, perhaps as a result of violent transients or noise.

These two measures are averaged over points of both bias conditions $z^{(1)}$ and $z^{(2)}$ corresponding to the alternate positive and negative going transitions. The total variance v_T is obtained from the following calculation:

$$v_T^2 = \sum_{i=1}^2 \left[\frac{1}{N_i - 1} \sum_{j=1}^{N_i} z_j^{(i)2} + \frac{1}{N_i (N_i - 1)} \left(\sum_{j=1}^{N_i} z_j^{(i)} \right)^2 \right] \quad V-14$$

Intrinsic variance (v_I) is estimated by:

$$\begin{aligned} v_I^2 &= \sum_{i=1}^2 \frac{1}{N_i} \sum_{j=1}^{N_i/2} (z_{2k}^{(i)} - z_{2k-1}^{(i)})^2 \\ &\approx \sum_{i=1}^2 \frac{1}{2N_i} \sum_{j=2}^{N_i} (z_j^{(i)} - z_{j-1}^{(i)})^2 \end{aligned} \quad V-15$$

where $z_j^{(i)}$ refers to the j th point of the set of either positive or negative going transitions.

These two terms may be particularly useful as a measure of signal degradation. They may be used as a measure of non-parametric correlation since the fraction of total variance "explained" by the intrinsic variance is given by:

$$A(z) = 1 - \frac{v_I^2}{v_T^2} \quad V-16$$

One estimate of non-parametric correlation is:

$$B(z) = \sqrt{A(z)} \quad . \quad \text{V-17}$$

Small values of $B(z)$ suggest that the apparent behavior of the raster pattern is not well characterized by a relatively smooth function of time.

3. Polynomial Mean Square Fits

In the hope of quantifying raster pattern behavior which system operators had described as "curve up," "curve down," "flat," or "complex," a fourth order polynomial mean square fit to the minimal phase difference raster data over the entire signal period was attempted for each bias condition. The fourth order fit was initially believed to be capable of representing both transient and steady state phenomena. Unfortunately, even the minimal phase difference representation of the data could not eliminate aliasing problems in those displays where apparently large initial transients occurred. As an alternative, a second order fit over that data lying outside the region of worst transient behavior was eventually incorporated into the measurement program.

The coefficients of the second order fit also serve the purpose of reducing the sensitivity of the Fourier transform data to phase offset and endpoint discontinuities which result from minor adjustments in the nominal raster period of the display.

As an example, consider the periodic Fourier series transform of a pattern characterized by a line with non-zero slope. As a result of endpoint discontinuity in a periodic representation, the odd Fourier series components exhibit amplitudes proportional to the magnitude of the discontinuity and decreasing in frequency as $\frac{1}{n}$.

It has previously been established that transient phenomena, zero point offset, and differences between the nominal and actual steady state period can produce patterns with both offset and non-zero slopes. If the raster pattern shows significant curvature, it is difficult to justify any particular choice for the nominal period. Any choice of nominal period creates its own contributions to the Fourier series components. Since the nominal raster period of the display is generally the result of an automatic measurement procedure, and since it represents one of the signal features currently used by the system's identification algorithm, the interests of comparability were best served by partially correcting for these discontinuities in terms of the mean square fit coefficients. Only the difference between the raster pattern and the polynomial mean square fit to each bias condition was used as input to the fast Fourier transform calculation.

4. Fourier Power Spectrum

In order to observe the effects of possible unintentional external modulation and periodic variation of the clock

rate, the use of Fourier power spectrum components was proposed. The calculation of these terms presented an intriguing problem, since the minimal phase raster data represent non-uniform time samples of a presumably continuous signal. The non-uniformity of sampling time arises from both the clock rate variation and the fact that the transition times themselves and data modulated. The presence of one transition per baud is not the case; transitions occur at various integer numbers of originator clock cycles.

The problem of non-uniform sampling is not too serious if one intends to obtain only a few Fourier power series components. The techniques for estimating amplitude phase and frequency of the principal Fourier components has been a topic in recent literature [22]. One can easily construct Fourier series of terms which are orthogonal on the interval $[0,L]$ by considering only the frequency terms which are periodic in L .

$$z(t) = \sum_{n=-\infty}^{\infty} \bar{c}_n e^{i \frac{2\pi n}{L} t} \quad \text{V-18}$$

These complex components

$$c_n = \int_0^L z(t) e^{-i \frac{2\pi n}{L} t} dt \quad \text{V-19}$$

may be approximated step-wise

$$c_n \approx \sum_{i=2}^{n_{\max}} z(t_i) e^{-i \frac{2\pi n}{t_{n_{\max}}} t_i} (t_i - t_{i-1}) \quad \text{V-20}$$

or by numerical integration involving higher order interpolation. Although numerical integration of the above type is possible, all frequency components up to the nominal clock frequency were potentially of interest as features. An alternate form of calculating the magnitudes of the power spectral components, using the fast Fourier transform was adopted for computer implementation.

For each bias condition, this procedure typically used about 75 non-uniformly spaced raster data points obtained outside the transient region. For each condition, 128 sample points uniformly spaced over the same time interval were chosen. The raster phase at these points was obtained by an interpolation procedure which performed a cubic polynomial fit to four minimal phase difference raster points, two on either side of the interpolation point. Since the fast Fourier transform routine used accepts complex values as input, the interpolated functions for positive going and negative going transitions were stored as the respective real and imaginary parts of this input. The complex input function may be expressed as:

$$y = f + ig \quad \text{V-21}$$

where the n th Fourier component of the real functions f and g are:

$$\begin{aligned} F_n(f) &= a_1 + i b_1 \\ F_n(g) &= a_2 + i b_2 \end{aligned} \quad \text{V-22}$$

The linearity of the Fourier transform and the requirement that for a real function

$$F_{-n}(f) = F_n^*(f) \quad \text{V-23}$$

implies that the n th Fourier transform components of both functions f and g may be recovered from $(F_n(y) \text{ and } F_{-n}(y))$, the real and imaginary Fourier components of the complex input [10]. Additionally it can be shown that:

$$\begin{aligned} &[\text{Re}[F_n(y)]]^2 + [\text{Im}[F_n(y)]]^2 + [\text{Re}[F_{-n}(y)]]^2 + [\text{Im}[F_{-n}(y)]]^2 \\ &= 2(a_1^2 + b_1^2 + a_2^2 + b_2^2) \end{aligned} \quad \text{V-24}$$

The Fourier transform information used as features in this thesis was the squared magnitude of the frequency components summed over both bias conditions in the manner shown above.

The Figures 12-22 show the behavior of the time quantization interval, the minimal phase difference interpolation, the second order polynomial fit, and the Fourier transform

features on a signal whose possible transition times t_n are given by

$$t_n = (nT + A \sin 2\pi f nT + B \exp(-anT)) \quad V-25$$

The data points were selected from these possible transition times by a simulated data modulation which repeated itself about $3\frac{1}{2}$ times in the duration of the display. In the raster patterns the values of the second order polynomial mean square fit and the interpolated raster phase for one bias condition at the 128 points used for interpolation are shown as the characters (+) and (*) respectively.

The effect of the quantization interval is apparent in all raster displays, but it is encouraging to note that a 25 Hz frequency component of amplitude equal to one-half the quantization interval (Figure 20) may be detected even in the presence of a transient during the initial half second of the signal sample. This detection ability, however, is strongly dependent on the apparently uniform distribution of transition times within the quantization interval. For this deterministic case, as the raster period approaches an integer multiple of the quantization interval, the ability of the measurement technique to isolate a signal of small amplitude as a single maxima deteriorates.

5. Laguerre Polynomial Coefficients

In the attempt to characterize the apparent transient behavior of some of the raster displays, a set of Laguerre

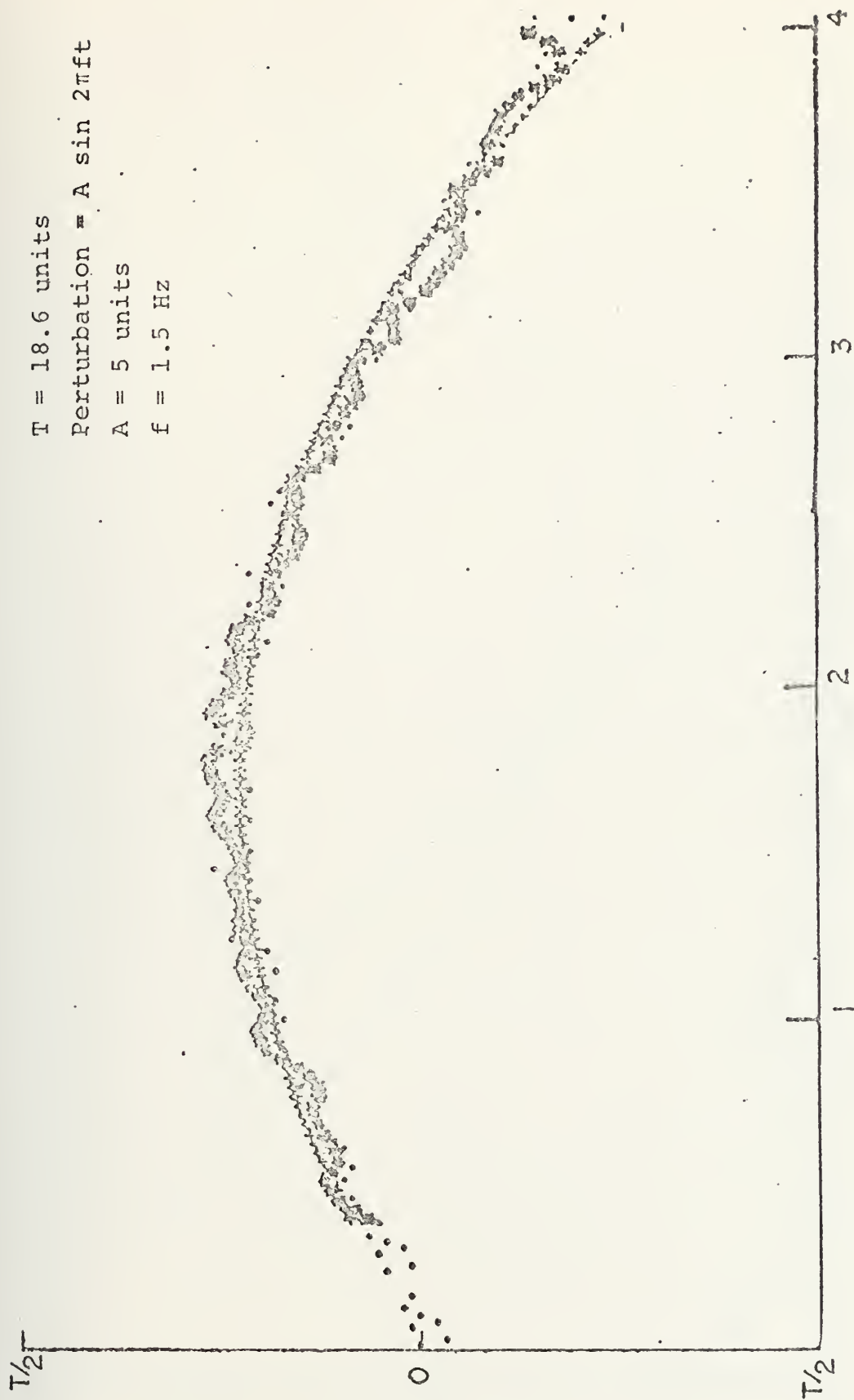


FIGURE 12. Polynomial Mean Square Fit and Interpolation, Signal 1



FIGURE 13. Fourier Magnitude Components - Signal 1

$T = 18.6$ units
 Perturbation = $A \sin 2\pi ft$
 $A = 5$ units
 $f = 3$ Hz

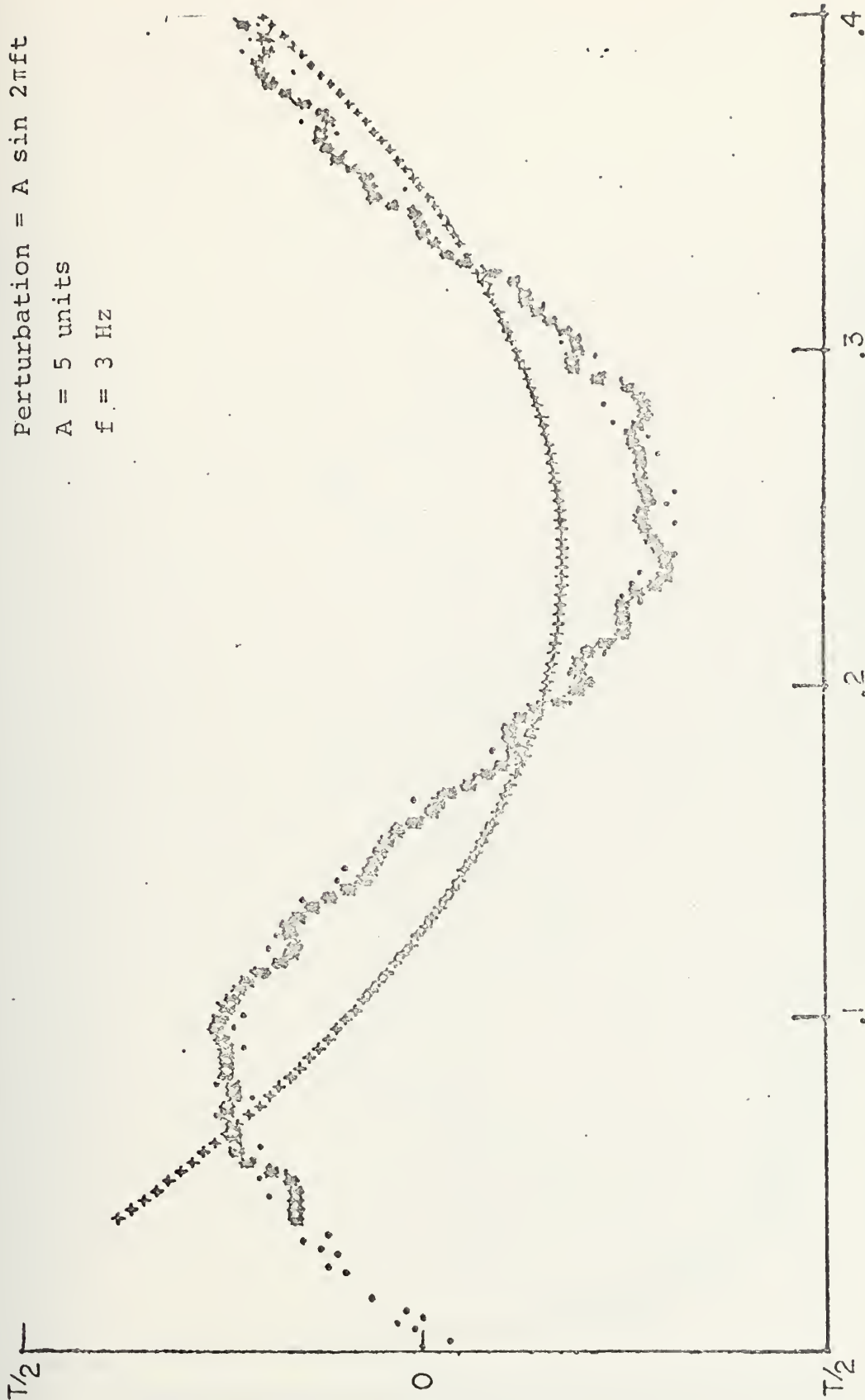


FIGURE 14. Polynomial Mean Square Fit and Interpolation, Signal 2

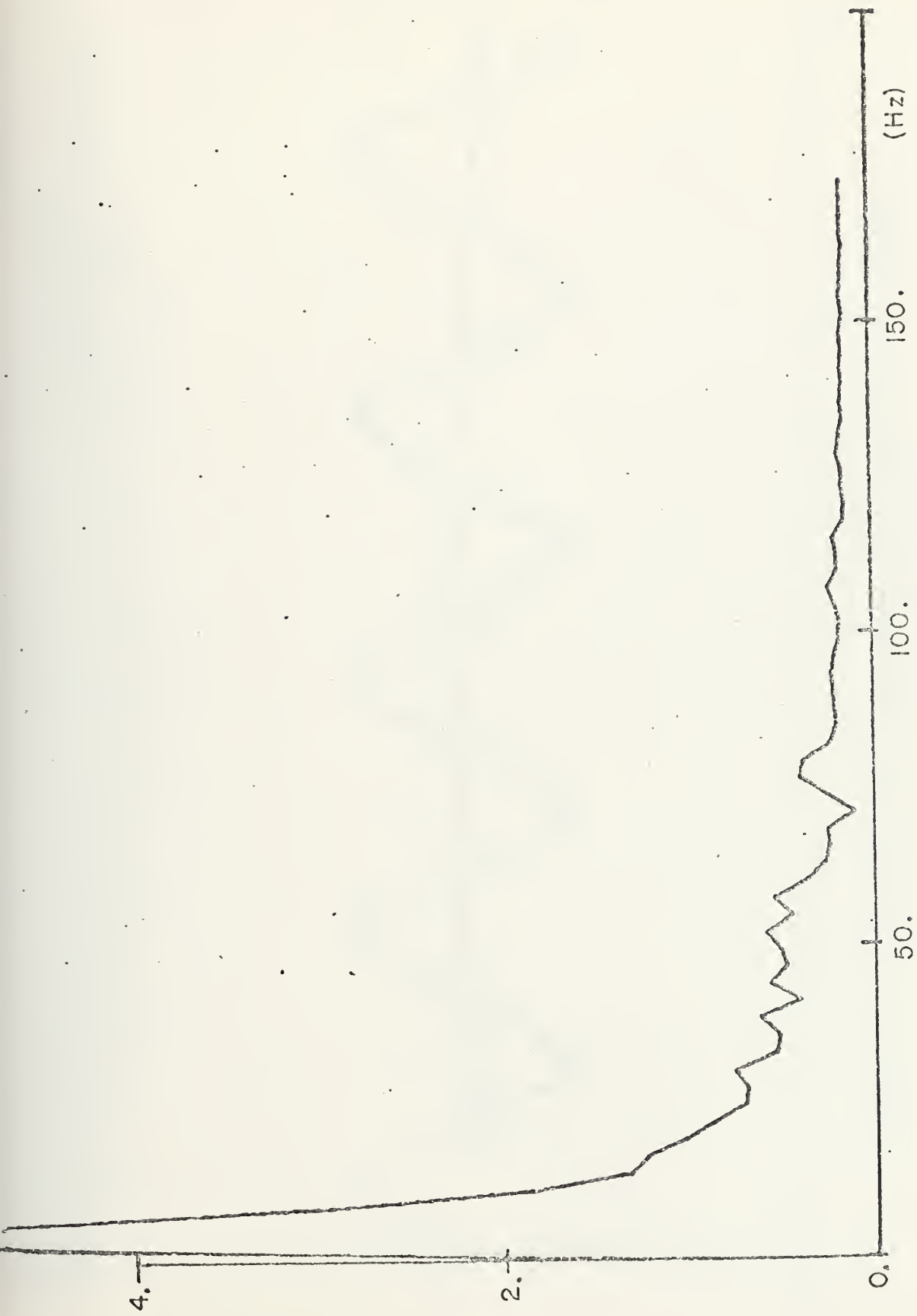


FIGURE 15. Fourier Magnitude Components - Signal 2

$T = 18.6$ units
 Perturbation = $A \sin 2\pi ft$
 $A = 2.5$ units
 $f = 12$ Hz

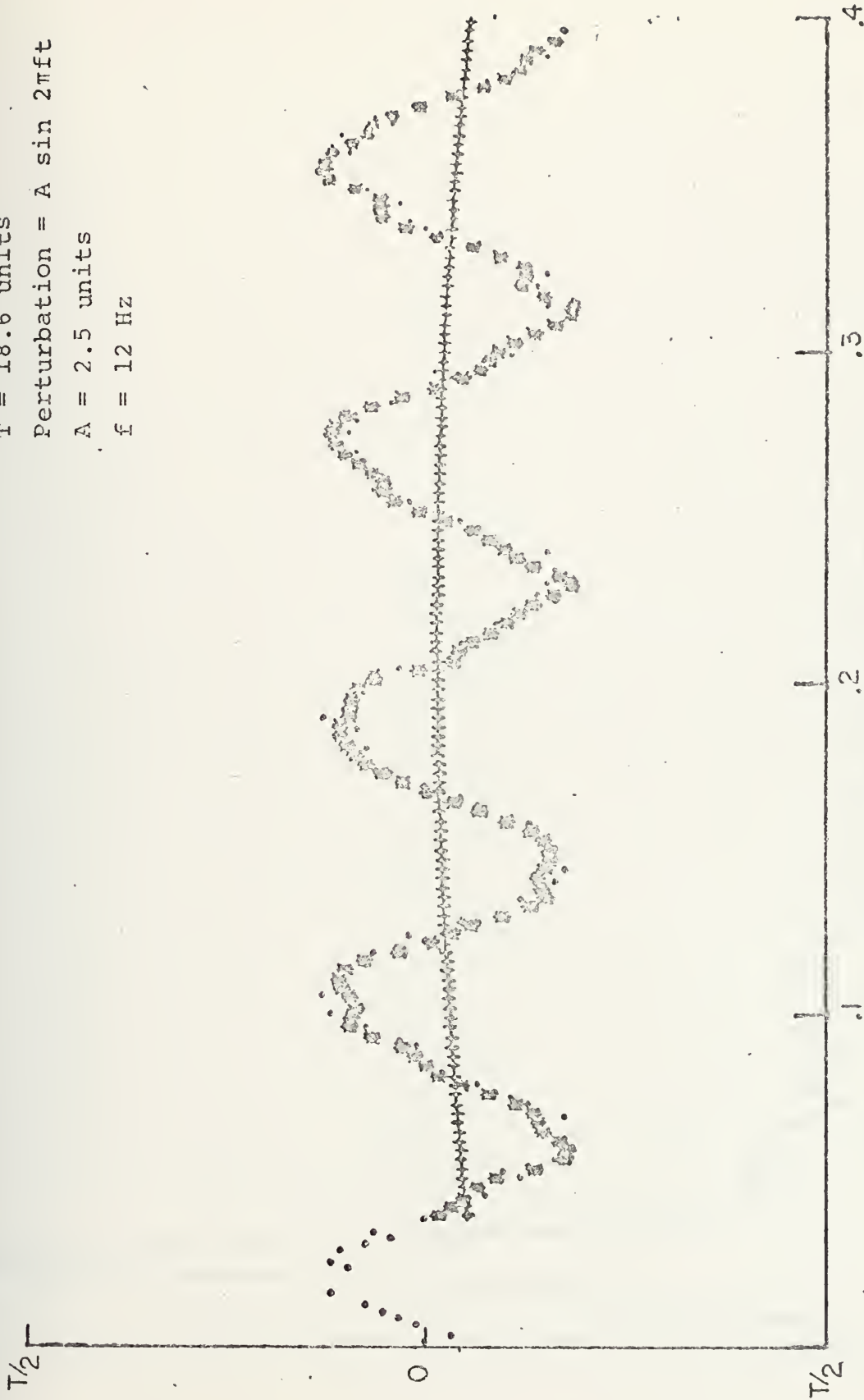


FIGURE 16. Polynomial Mean Square Fit and Interpolation - Signal 3

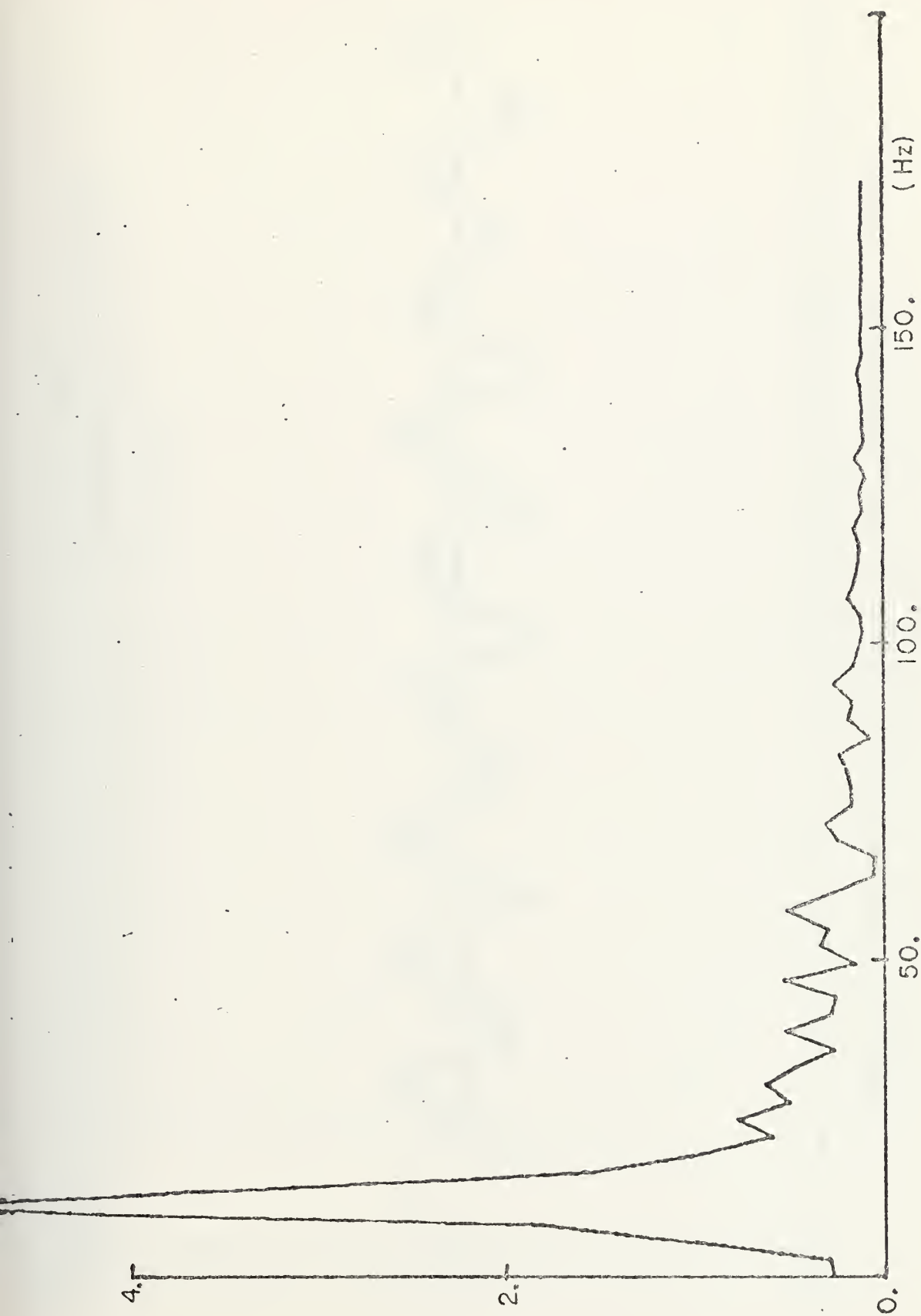


FIGURE 17. Fourier Magnitude Components, Signal 3

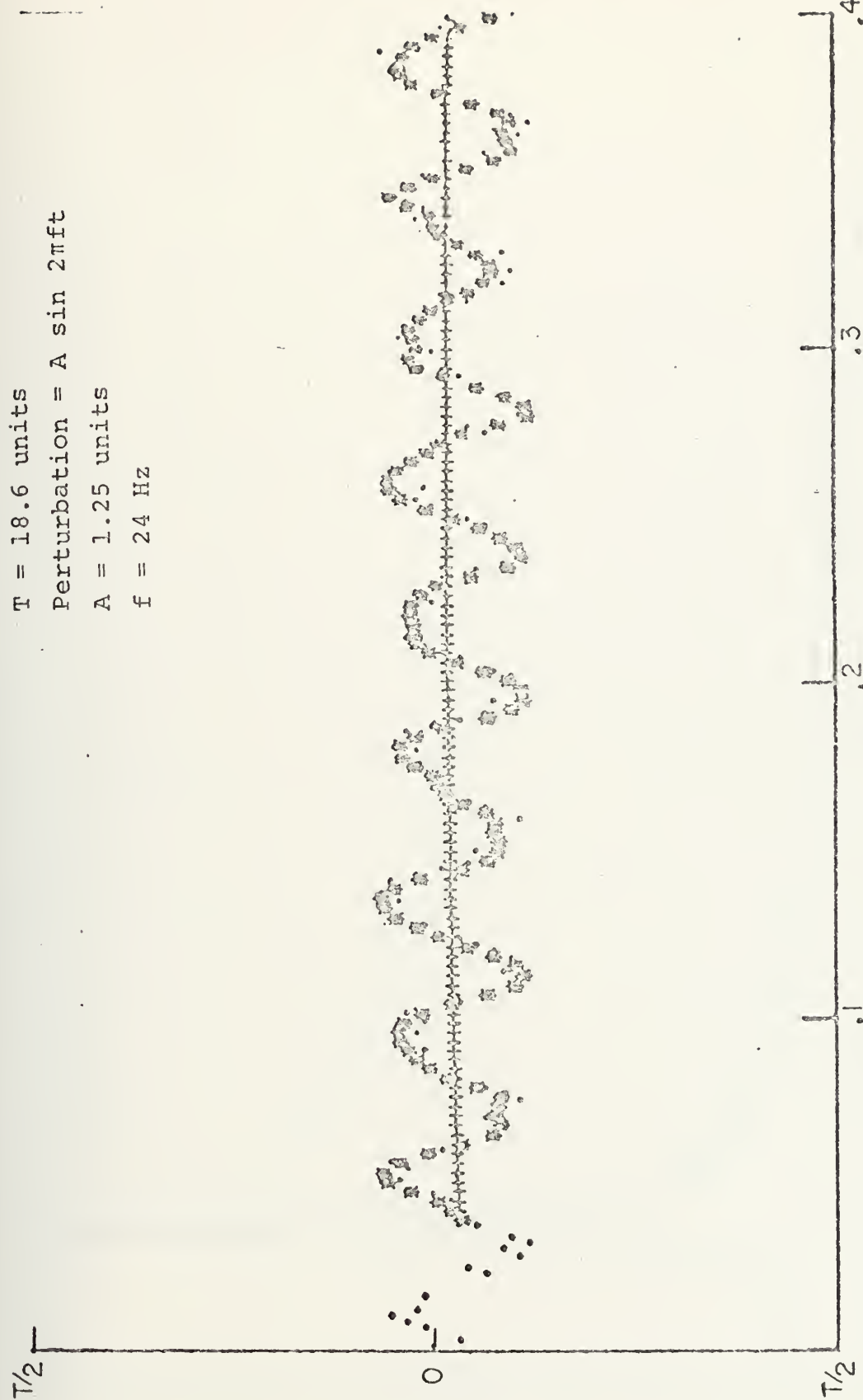


FIGURE 18. Polynomial Mean Square Fit and Interpolation, Signal 4

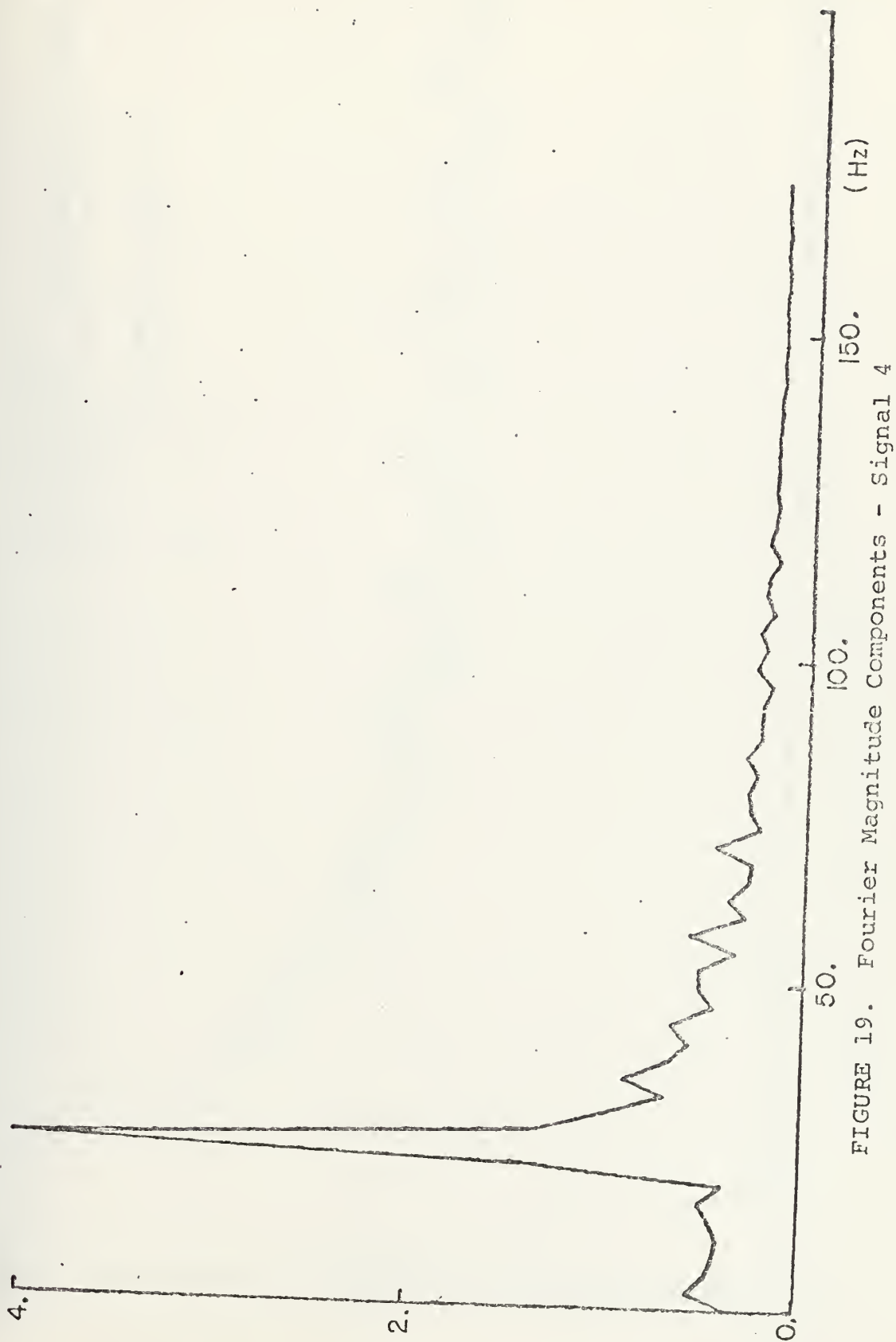


FIGURE 19. Fourier Magnitude Components - Signal 4

$T = 18.6$ units
 Perturbation = $A \sin(2\pi ft) + B \exp(-at)$
 $A = .5$ quantization intervals
 $f = 25$ Hz
 $B = 20$ quantization intervals
 ≈ 1 raster period
 $\frac{1}{a} = .05$ sec.

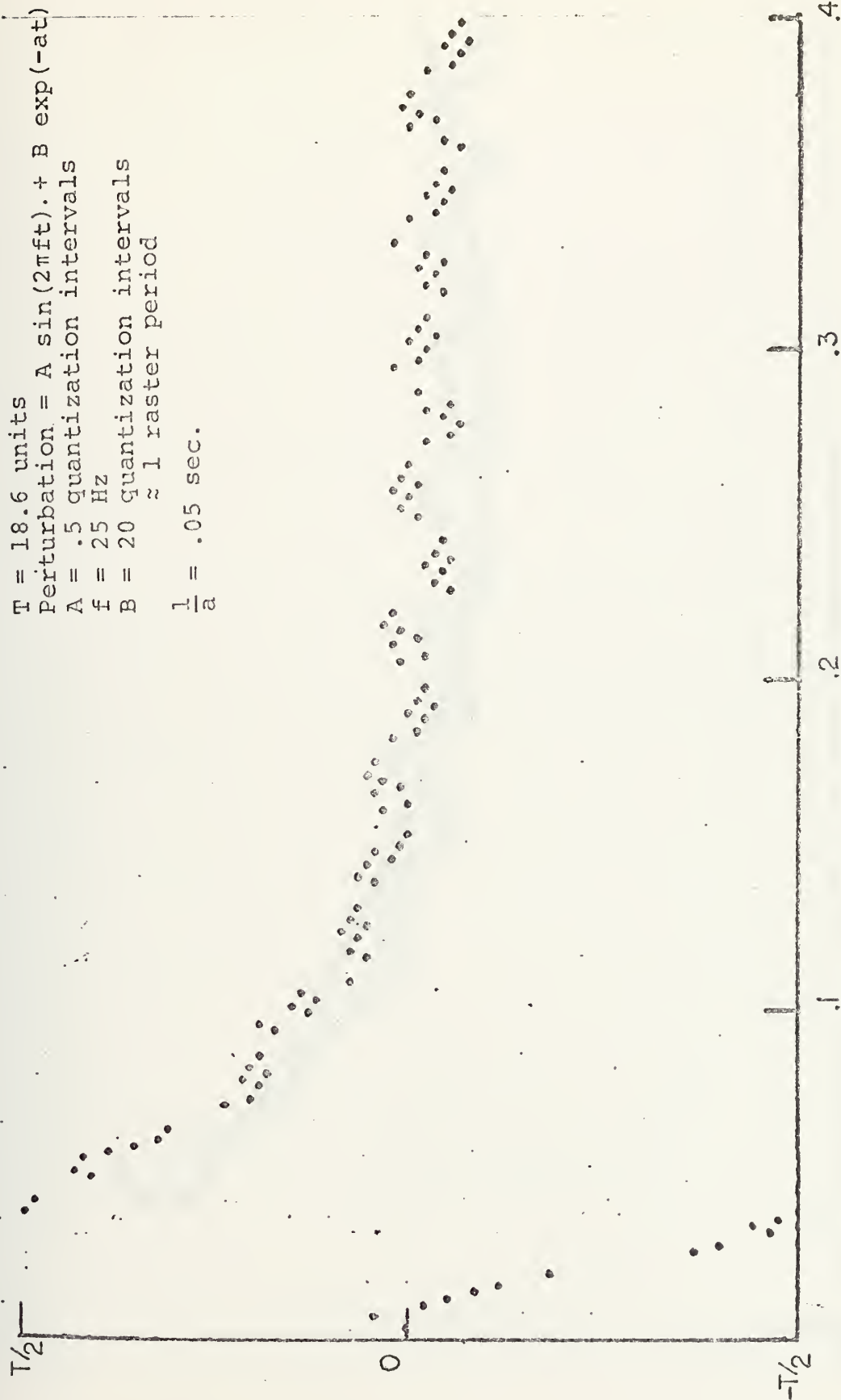


FIGURE 20. Raster Display, Signal 5

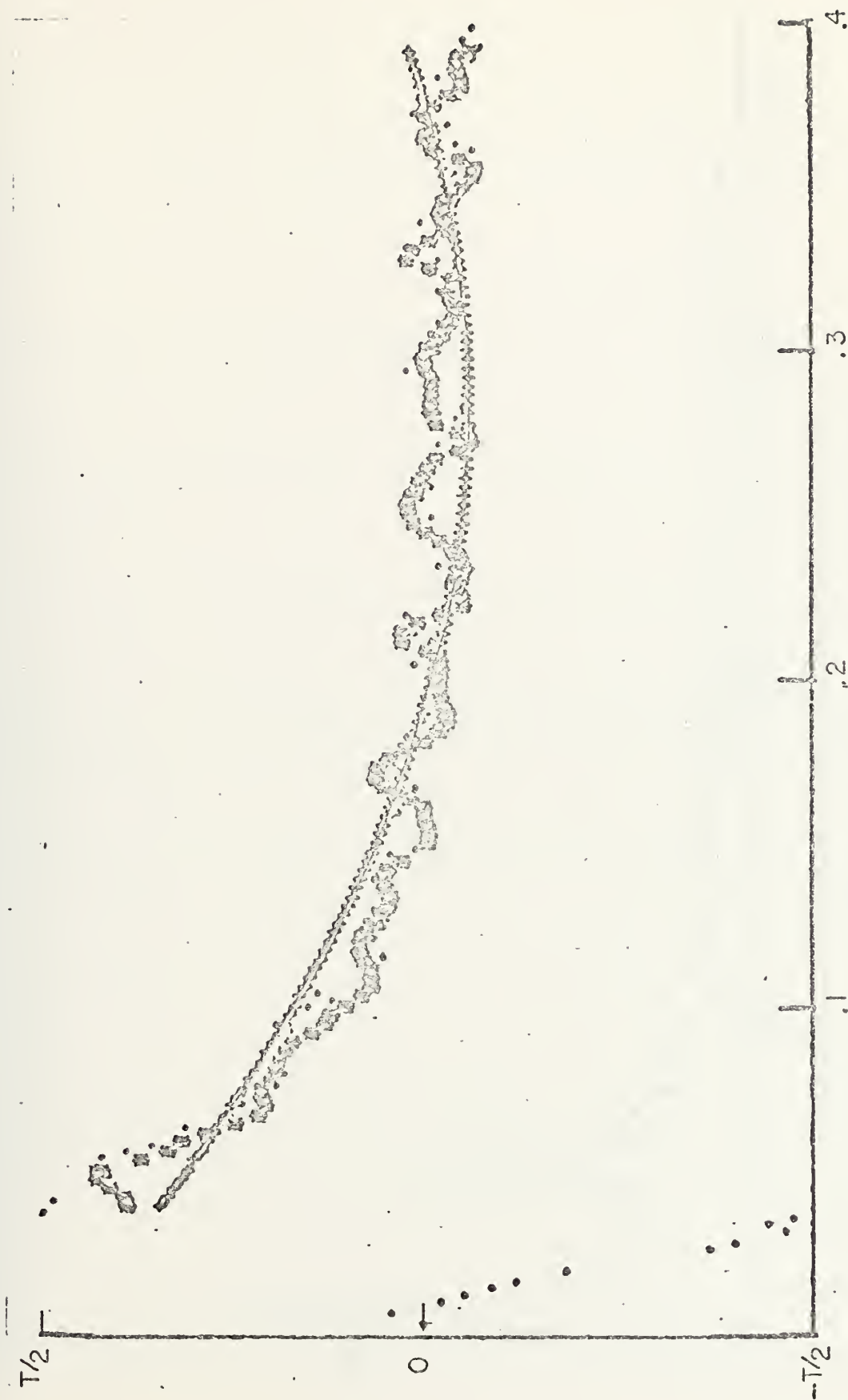


FIGURE 21. Polynomial Mean Square Fit and Interpolation, Signal 5

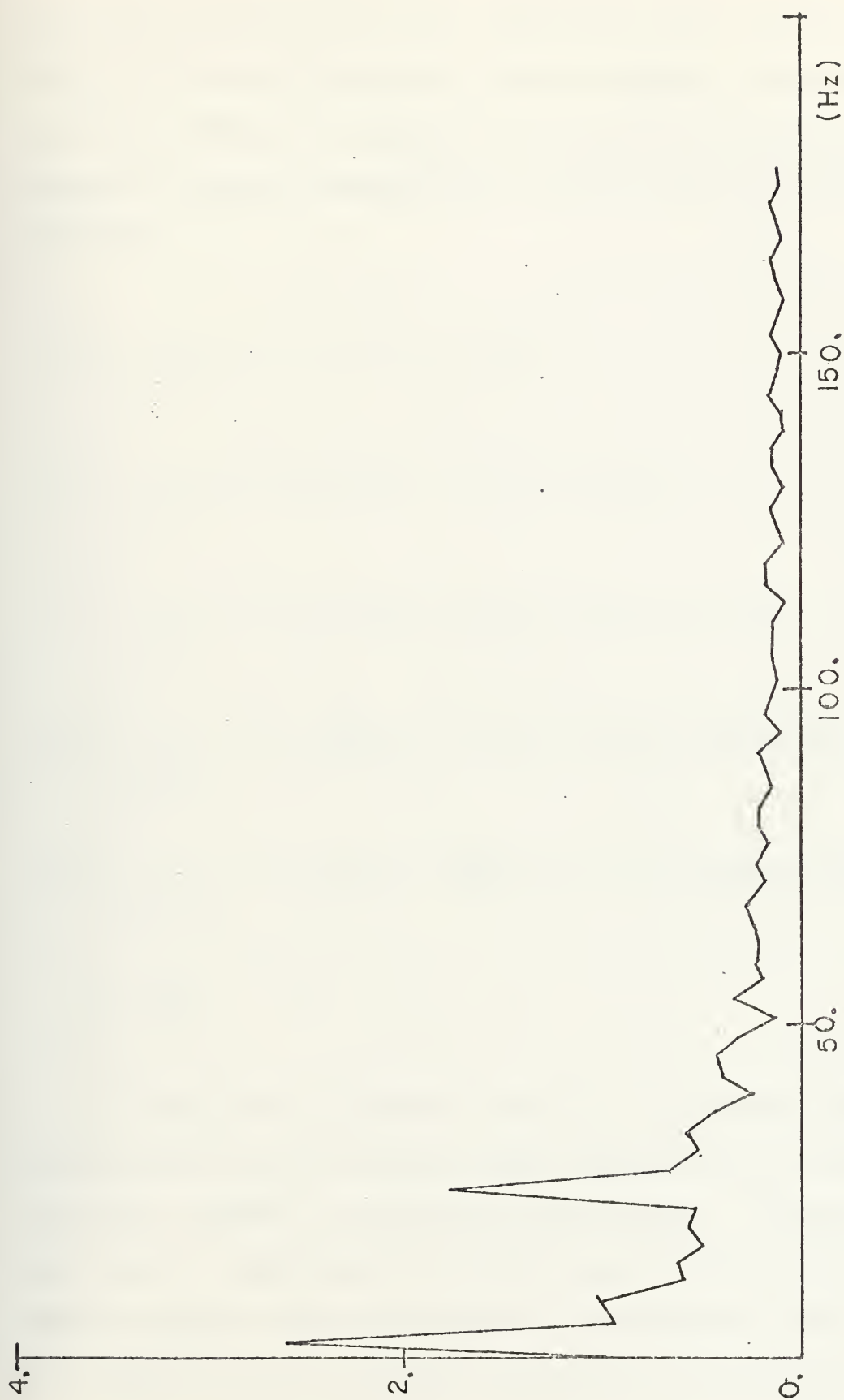


FIGURE 22. Fourier Magnitude Components, Signal 5

polynomial coefficients were calculated over all the points $z'(t_i)$ of the initial portion of the signal. The coefficients L_n of the first five Laguerre polynomials were obtained by stepwise numerical integration over the first 800 quantization intervals.

$$L_0 = \sum_{t_i \leq 800} z'(t_i) \exp[-\rho t_i] (\Delta t_i)$$

$$L_1 = \sum_{t_i \leq 800} z'(t_i) [2\rho t_i - 1] \exp[-\rho t_i] (\Delta t_i)$$

$$L_2 = \sum_{t_i \leq 800} z'(t_i) [2(\rho t_i)^2 - 4\rho t_i + 1] \exp[-\rho t_i] (\Delta t_i)$$

$$L_3 = \sum_{t_i \leq 800} z'(t_i) \left[\frac{4}{3}(\rho t_i)^3 - [6\rho t_i]^2 + 6\rho t_i - 1 \right] \exp[-\rho t_i] (\Delta t_i)$$

$$L_4 = \sum_{t_i \leq 800} z'(t_i) \left[\frac{2}{3}(\rho t_i)^4 - \frac{16}{3}(\rho t_i)^3 + 12\rho t_i - 8\rho t_i + 1 \right] \exp[-\rho t_i] (\Delta t_i)$$

V-26

$$\rho = \frac{1}{80}, \quad \Delta t_i = t_{i+1} - t_i$$

During the transient period, the aliasing effect, which occurs if the nominal raster period does not match the actual period, is especially pronounced. The data may very well be undersampled in this region, thus no attempt was made to compute these coefficients independently for each bias condition and the phase unwrapping procedure was applied

to all data points. This process, however, operated on the premise that the phase difference from transition to transition is not more than one-half the nominal raster phase. Even if both positive and negative transitions are used for phase unwrapping, this premise may not be valid.

If the initial transition times predicted by the steady state period and knowledge of the signal's underlying data modulation can be determined, a more accurate representation of transient phase might be obtained. Unfortunately, knowledge of this data modulation was not available for the purpose of this thesis.

D. COMPUTER FILES AND PROGRAMS

The programs used to produce the feature measurements described in the previous section were designed to operate on a pre-production model of the Parameter Encoder which incorporated in its hardware a Hewlett-Packard 2100A computer with twin disc drives and moving head disc operating system. The normal graphics display and list device was a Tektronix storage scope.

The programs discussed in this section, while written in HP Fortran, make use of the Fortran-callable executive routines available under the disc operating system. These routines permit program overlays and disc file input/output. Similarly, a library of Fortran-callable utility routines developed at Electromagnetics Systems Laboratory to control graphics in put and output were used to create and erase the

scope displays, temporarily halt computations and provide keyboard control of the measurement process.

The signal data base used as input to the measurement process was stored in the disc file, PARF. This file contained a signal index number, the automatically measured clock rate, the time difference between successive signal transitions and the total number of transition times measured (see Table 1). Since this data format differed slightly from that normally accepted by the standard raster display routine RASF, an interface program FACE was written to convert the data.

Following the initial display of the raster signal, process control could be transferred to a program FOURD whose function was to call the interpolation, measurement, and display subroutines and to store the measured parameters in the disc file PAR. The overall process flow of FACE, RASF and FOURD is shown in Figures 23 and 24.

1. Measurement Process

The first subroutine called by FOURD is RESDU (see Figures 25 and 26 for overall process flow). This subroutine calculates a minimal phase difference representation for the points of both bias conditions outside the transient region. The points thus represented are used by the subroutine CURV to calculate the coefficients of the polynomial mean square fit. They are also used to calculate the signal's mean bias and its total and intrinsic variance. A cubic

TABLE I
Disc File PARF

Size: 1542 sectors - 128 Integer words per sector

Sectors 0 - 5 Directory - 768 words

<u>Word</u>	<u>Use</u>	<u>Comment</u>
1-768	Signal number	Order in directory corresponds to order of signal files

Sectors 6 - 1542 Signal Files 2 Sectors per signal

<u>Word</u>	<u>Use</u>	<u>Comment</u>
1 - 2	Nominal period	Floating point
3	Number of measurements	
4 - 256	Transition time	Time difference between successive transitions

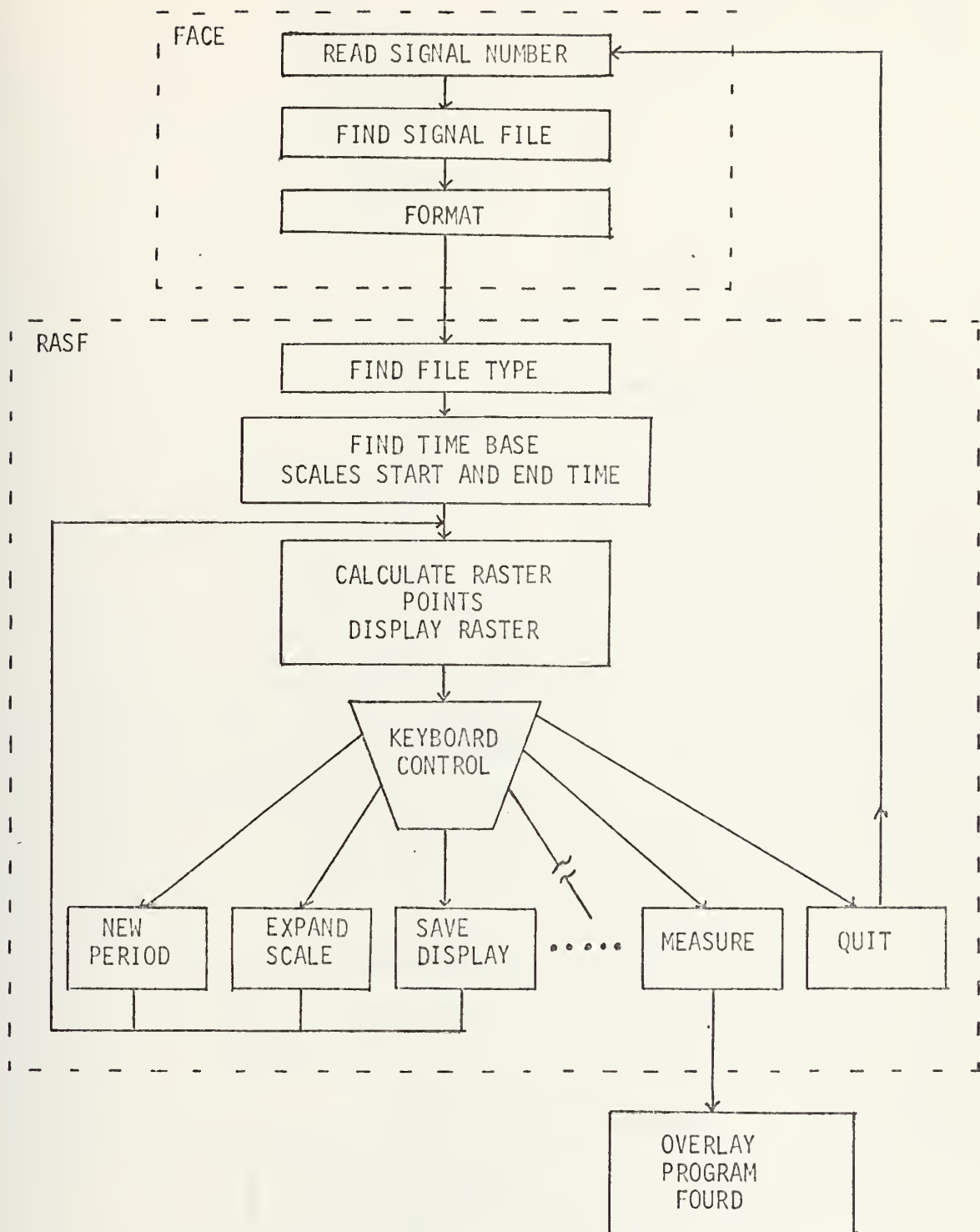


FIGURE 23. Process Flow, Interface and Raster Display Programs FACE and RASF

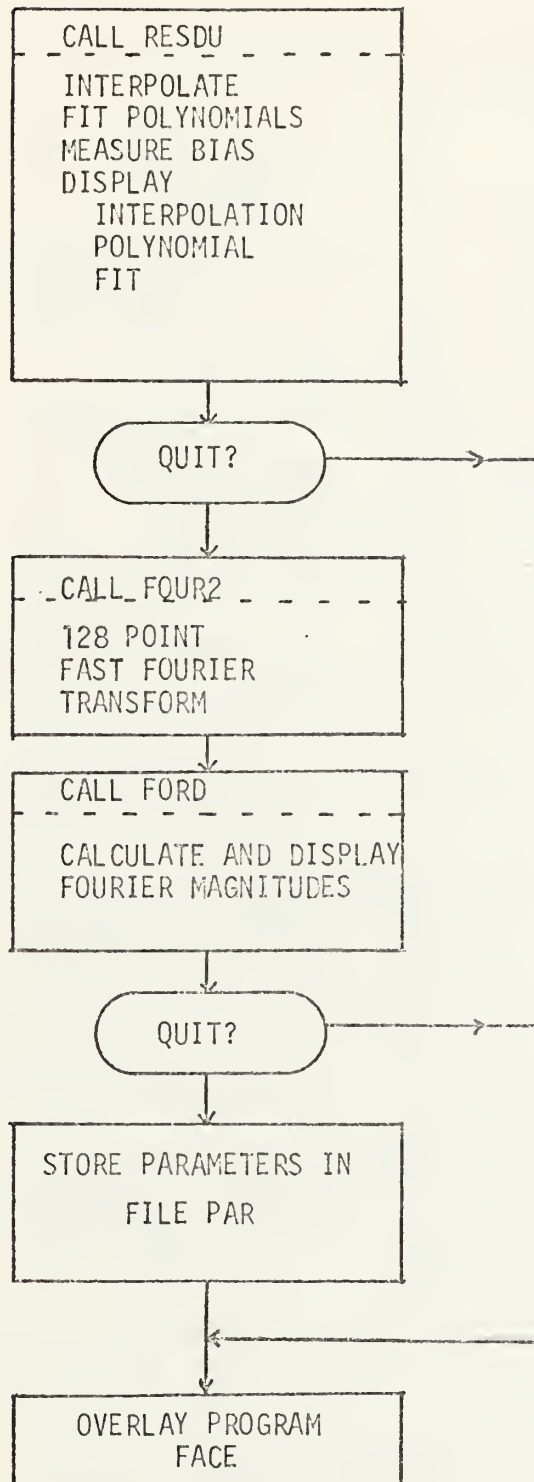


FIGURE 24. Process Flow Measurement and Display Program FOURD

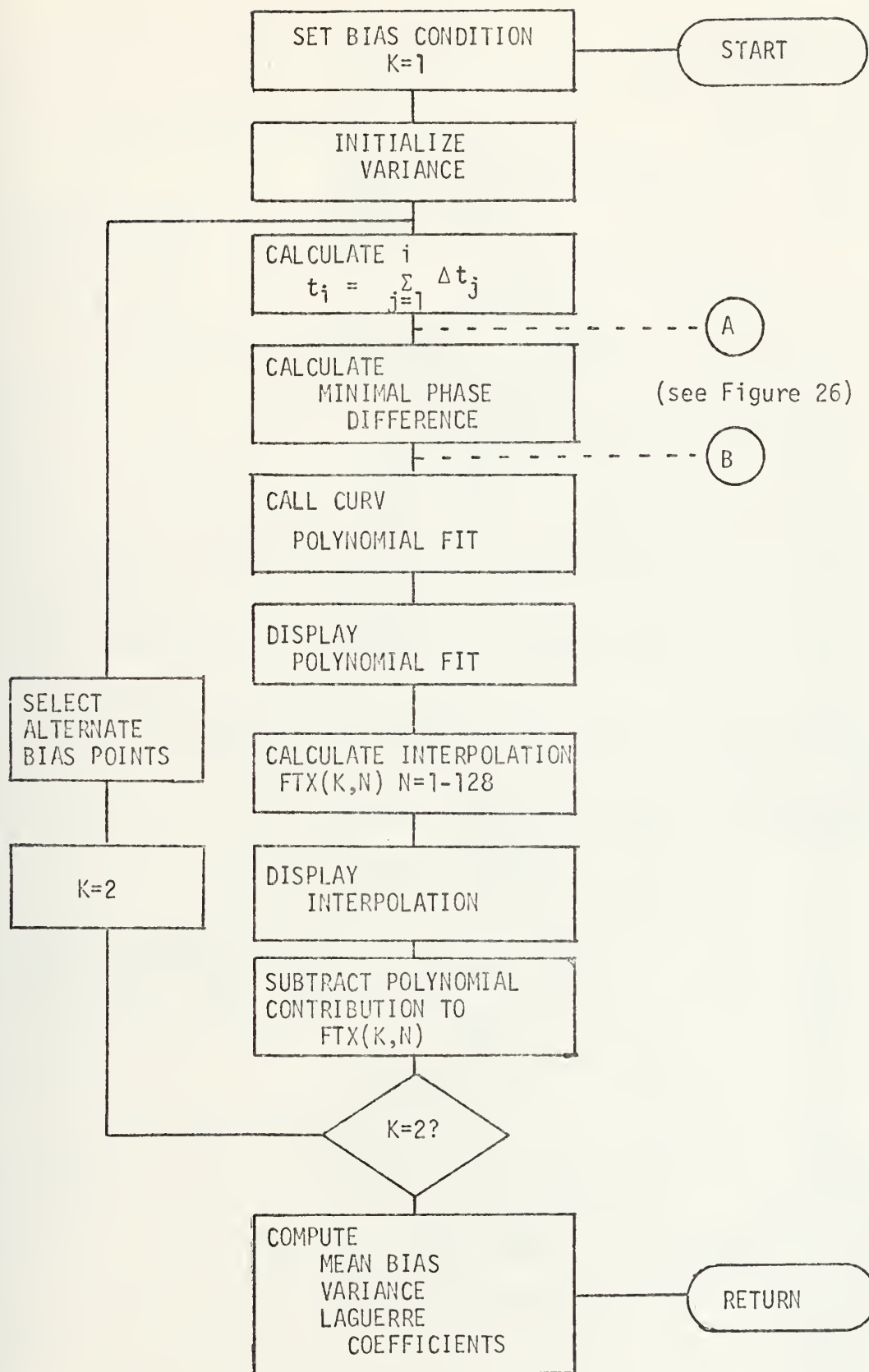


FIGURE 25. Flow Diagram Interpolation and Measurement Subroutine RESDU

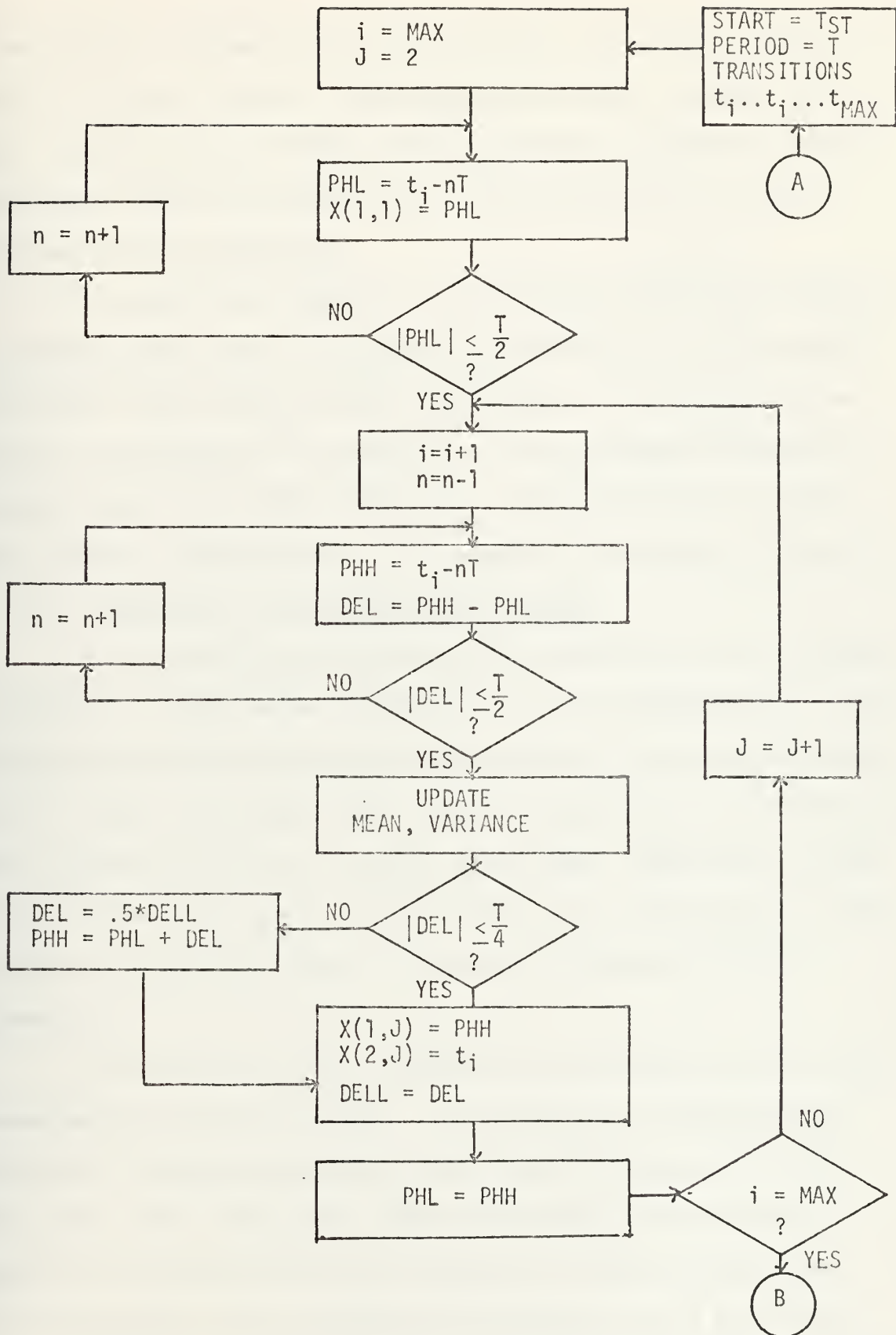


FIGURE 26. Minimal Phase Difference Algorithm

interpolation scheme CURF then provides the uniform samples used by the fast Fourier transform subroutine. RESDU also calls the subroutine LAGER, which performs a stepwise numerical integration to estimate the coefficients of the first five Laguerre polynomials.

The two other subroutines called by FOURD are FOUR2, a standard fast Fourier transform, and FORD, a routine which calculates the average square magnitude of the Fourier components and displays the results on the Parameter Encoder's storage scope. The disc file, PAR, which contains all the above feature measurements, is organized as shown in Table II.

2. Statistics and Formatting Routines

The analysis of the features measured by the routines indicated in FOURD makes extensive use of identification and analysis software previously developed for the Parameter Encoder. These procedures use a yet another disc file structure, MASTF (Table III), as their input data base. Additional reformatting and selection of the parameters was necessary to allow the use of this file since it accepts a maximum of 50 parameters.

The program CSTAT provided class statistics for the measurements stored in MASTF. In addition to reformatting individual signal parameters, and storing them in the disc file MASTF, the subroutine TRNSG performed the important function of interpolating the existent Fourier magnitudes to estimate the square magnitude components at 2 Hz increments. This was necessary since the resolution of each signal

TABLE II

Disc File PAR

Size: 200 Sectors (128 Integer Words per Sector)

Sector 0 Directory

<u>Word</u>	<u>Use</u>	<u>Comment</u>
1-128	Signal Number	Order in directory corresponds to order of files

Sectors 1-199 Signal Parameter Files 2 Sectors per File

<u>Word</u>	<u>Use</u>	<u>Comment</u>
1-2	Nominal Period	Floating Point SIG(1)
3-4	Signal Resolution	Floating Point SIG(2)
5-8	(Not Used)	
9-18	Polynomial Coefficients Mean Square Fit	Bias Condition 1 SIG(5-9)
19-28	Polynomial Coefficients Mean Square Fit	Bias Condition 2 SIG(10-14)
29-38	Polynomial Coefficients Laguerre	Floating Point SIG(15-19)
39-40	Mean Bias Estimate	Floating Point SIG(20)
41-70	(Not Used)	
71-72	Variance	Floating Point SIG(35)
73-74	Intrinsic Variance	Floating Point SIG(36)
75-78	(Not Used)	
79-208	Fourier Magnitude Components	65 Terms SIG(40-104)
209-256	(Not Used)	

TABLE III

Disc File MASTF

Size: 1524 Sectors 128 Integer Words per Sector

Sector 0 File Information

<u>Word</u>	<u>Use</u>
1	Number of Sectors in MASTF
2	Number of Words in Directory
3	Number of Words in File
5	Scale Code (-1 No scaling)
6	Variance Breakpoint
14	Number of Parameters
15-64	Parameter Numbers

Signal File

<u>Word</u>	<u>Use</u>	<u>Comment</u>
1	Signal Number	
2-3	Identification Number	
4	(Not Used)	
5-104	Parameters	50 Floating Point Words
105-154	Parameter Weight	1 If used, 0 If Not Used
159-384	(Not Used)	

Class File

<u>Word</u>	<u>Use</u>	<u>Comment</u>
1-2	Identification	
3	Number of Entries	
4	Number of Parameters	
5-104	Class Means	Floating Point
105-204	Class Variance	Floating Point
205-254	Number of Parameters Used for Classification	
255-304	Parameter Numbers	
305-384	(Not Used)	

was variable. Ground truth classification of each signal had been previously determined through the use of parameters not included in this thesis, and was stored in a small Disc File, CSIGF.

E. ANALYSIS OF PARAMETERS

From the basic group of features obtained by the measurement process, a group of 50 parameters was selected for further analysis. The analytic techniques used were those readily available to the user of Parameter Encoder identification software. Several iterations of measurement, observation, and feature analysis were conducted before arriving at the parameters indicated in Table IV. Earlier measurements had involved the use of fourth order polynomial mean square fits to the entire raster display, and the use of Fourier magnitude coefficients of raster pattern input uncorrected by the polynomial fit.

These earlier techniques appeared to provide reasonable measurements for approximately 60 percent of the signals considered, but failed catastrophically for the remaining signals. The major cause of failure appeared to be the inability of the minimal phase difference representation to track the raster pattern through initial transients and anomalous transition times. The final measurement process discussed in the previous section was applicable to about 80 percent of the signals attempted and resulted in measurements for the final training set of 74 signals distributed among nine classes as shown in Table V.

TABLE IV

Parameter Identification

<u>Parameter #</u>	<u>Identification</u>
901	Nominal Period
902	Length of Signal Resolution
903	0 Order Polynomial Coefficient Bias Cond. 2
904	1st Order Polynomial Coefficient Bias Cond. 2
905	2nd Order Polynomial Coefficient Bias Cond. 2
906	0 Order Polynomial Coefficient Bias Cond. 1
907	1st Order Polynomial Coefficient Bias Cond. 1
908	2nd Order Polynomial Coefficient Bias Cond. 1
909	1st Laguerre Coefficient
910	2nd Laguerre Coefficient
911	3rd Laguerre Coefficient
912	4th Laguerre Coefficient
913	5th Laguerre Coefficient
914	Mean Bias
915	Total Variance
916	Intrinsic Variance
917	Fraction of Total Variance Explained by Intrinsic
918	Fourier Power Component 0 Hz
919	Fourier Power Component 2 Hz
.	
.	
.	
950	Fourier Power Component 64 Hz

TABLE V

Training Set Signal Distribution

<u>Class 1</u>	(Not Used in Distance Ratio)			
1447	1300	875	126	171
172	775	264		
<u>Class 2</u>				
477	758	676		
<u>Class 3</u>				
770	183	192	1483	1488
693	1354	1367	1498	818
306	307	319	1525	938
<u>Class 4</u>				
1441	1444	483	1309	771
1327	1337	1338	742	315
1399	925			
<u>Class 5</u>				
127	155	766	826	
<u>Class 6</u>				
489	235	236	787	1543
<u>Class 7</u>				
1286	1289	672	475	674
675	120	1304	1305	141
<u>Class 8</u>	(Not Used in Distance Ratio)			
888	239	1487	698	915
446	970			
<u>Class 9</u>				
121	754	145	149	196
189	201	202	203	204

The ground truth identification of these signals was based on the agreement of two external identification schemes. One of these was the semi-automated process developed by Electromagnetic Systems Laboratory which uses additional signal parameters not associated with the raster display. The other process uses all-source information in arriving at signal identifications.

The analysis began with the measurement of single feature separating capability for each of the 50 selected parameters. The eleven features showing the greatest capability were then examined for redundant features by calculation of class and global correlation matrices. Finally, the error rate of a minimum distance classifier was used as the criteria for a feature search procedure which combined single feature ranking and search-without-replacement techniques. Since the amount of training data was limited, even this iterative procedure may place too great an emphasis on the training set values.

1. Single Feature Separability

The single feature separability measure available in the Parameter Encoder software is a modification of the distance ratio techniques discussed in Appendix C.2. The program FINTR calculates the average square distance between points of m different classes,

$$D_B^2 = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i}^m \frac{1}{K_i K_j} \sum_{k=1}^{K_i} \sum_{l=1}^{K_j} (x_l^{(i)} - x_k^{(j)})^2 \quad V-27$$

The m class average within-class-variance,

$$D_w = \frac{1}{m} \sum_{k=1}^m \frac{2}{K_k(K_k-1)} \sum_{i=1}^{K_k} \sum_{j=i}^{K_k} (x_i^{(k)} - x_j^{(k)})^2 \quad V-28$$

and uses the ratio

$$F = \frac{D_B^2}{D_w^2} \quad V-29$$

as a measure of separability. Since the array storage available in this program limits the number of input samples to 60, the signals of classes 1 and 8 were not used in the calculation of the feature distance ratios. The distance ratios obtained using all signals of the other 7 classes are shown in Table VI. The distance ratios of the frequency components are also plotted in Figure 27.

By way of comparison Figure 28 shows the distance ratio as a function of frequency for data obtained in earlier investigations using a smaller number of signals and classes. The well defined maxima in the latter case can be attributed to the fact that the signals and classes user were particularly noise and transient free. In these cases one was able to use the entire raster display uncorrected by a polynomial fit to estimate the Fourier components.

If one assumes that the parameters are statistically independent, then the ranking of features according to distance ratio provides a guide for feature selection. Eliminating

TABLE VI

Single Distance Ratios for Parameters 901 - 950

<u>#</u>	<u>Ratio</u>	<u>#</u>	<u>Ratio</u>	<u>#</u>	<u>Ratio</u>
901	10.59	921	1.32	941	1.14
902	11.68	922	1.14	942	1.08
903	1.87	923	1.35	943	.94
904	1.07	924	1.25	944	.93
905	1.07	925	1.45	945	.93
906	1.17	926	1.81	946	.84
907	1.00	927	2.10	947	.95
908	1.03	928	1.51	948	.92
909	1.15	929	1.09	949	.96
910	1.03	930	1.44	950	.97
911	1.01	931	1.68		
912	.99	932	1.06		
913	1.08	933	1.27		
914	.98	934	1.11		
915	.99	935	1.28		
916	1.20	936	1.44		
917	1.14	937	1.35		
918	1.68	938	1.26		
919	1.22	939	1.02		
920	1.19	940	1.01		

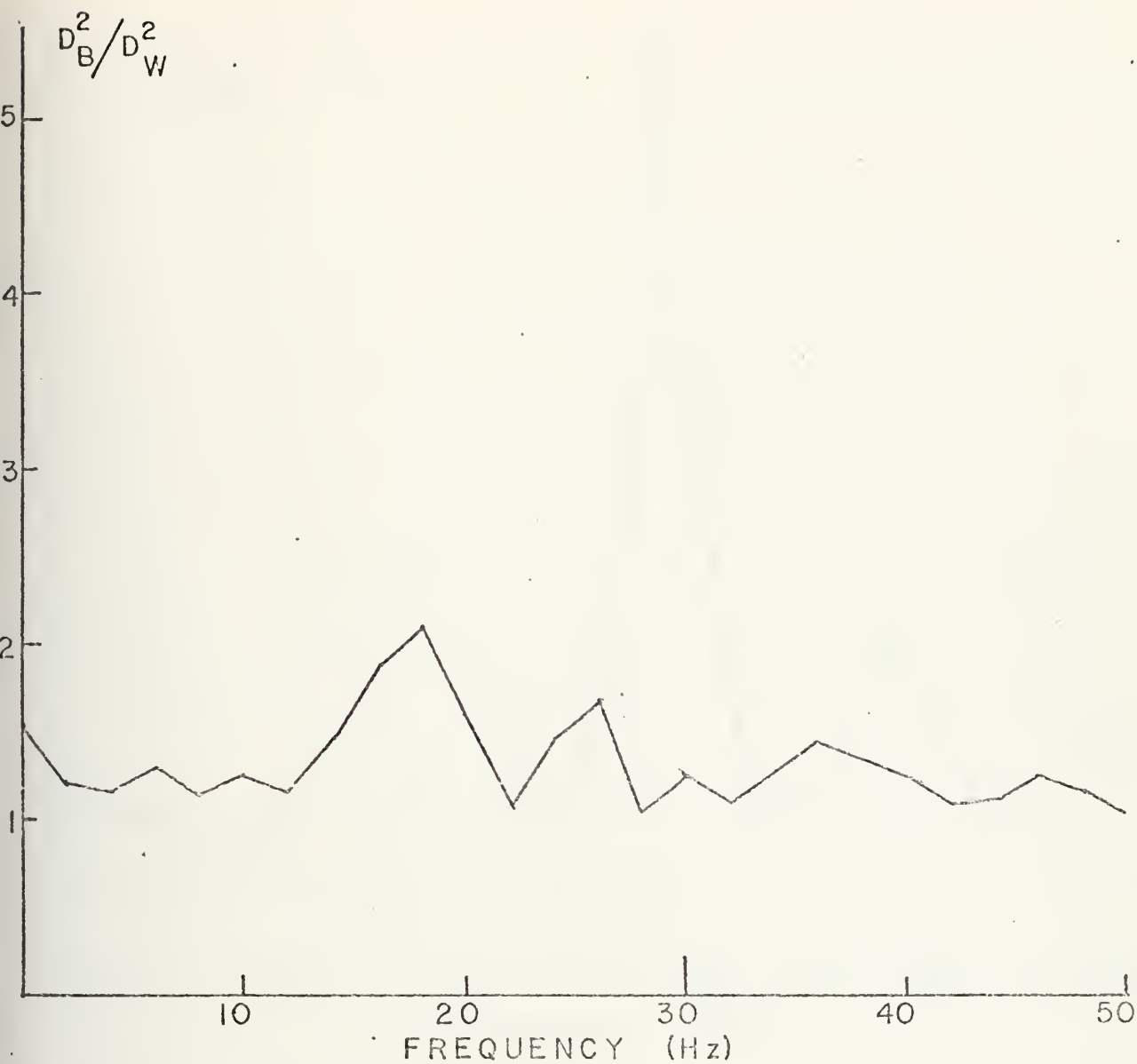


FIGURE 27. Feature Separation Quality vs. Frequency, 59 signals, 7 classes

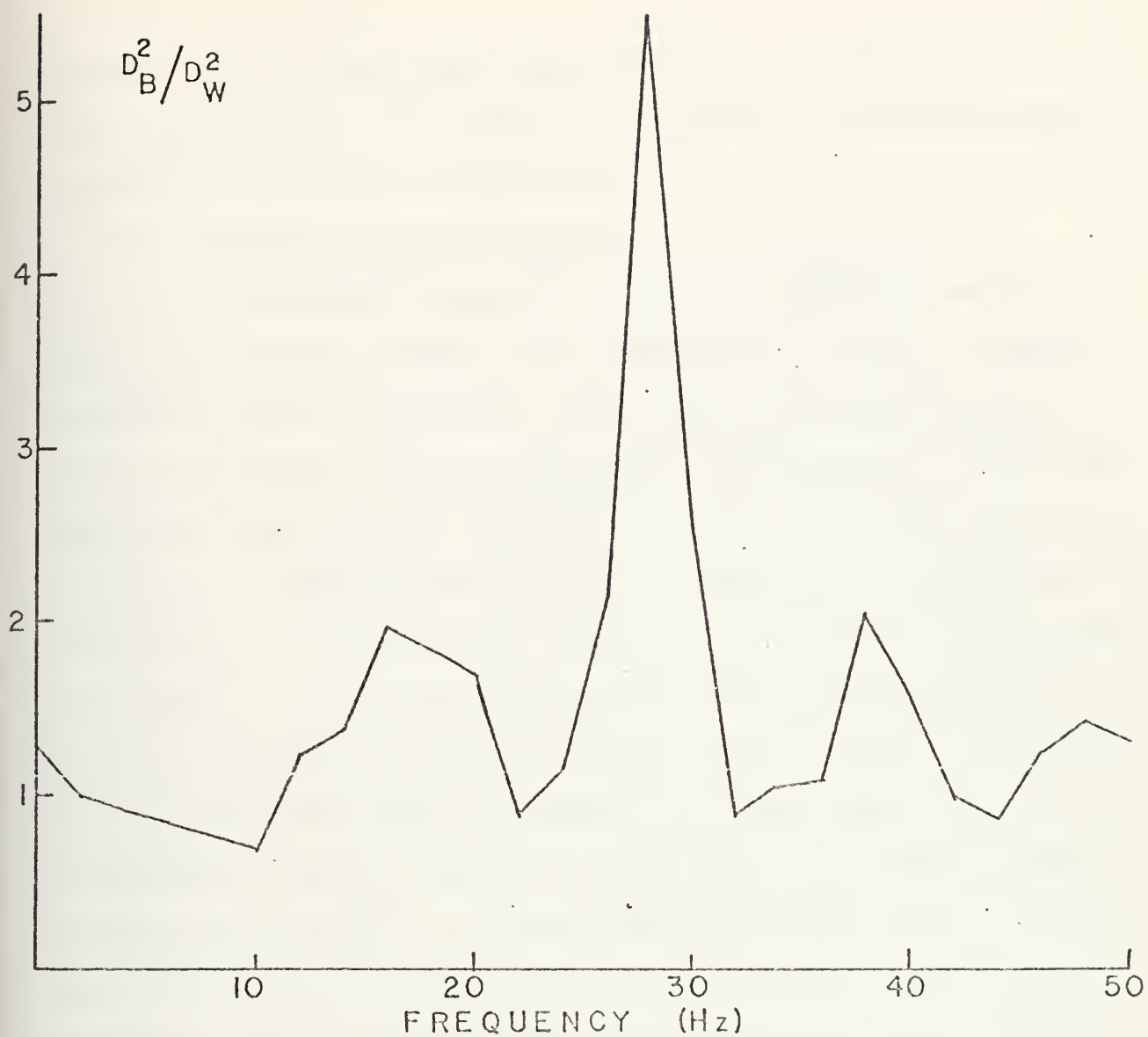


FIGURE 28. Feature Separation Quality vs. Frequency, 30 signals, 4 classes, no polynomial fit corrections

those features whose ratio was less than one and ranking the best eleven of the remaining resulted in the feature set shown in rank order in Table VII.

2. Multiple Feature Analysis

To test the assumption that the parameters selected by single feature ranking were independent a utility routine PCORL was used to calculate their class and global correlation coefficients. Two dimensional scattergrams for selected pairs of highly ranked features were used for visual confirmation of feature correlation and clustering. In the global case, two of the parameters exhibited correlation coefficients in excess of 0.9. These parameters, whose scattergram is shown in Figures 29 and 30, are the nominal raster period and the signal resolution. Correlation between these two parameters might be expected to be high if the sample signal length corresponds to the same number of bauds for all signals. Both the scattergram and the class correlation coefficients indicated that the period and signal duration remain highly correlated on a class basis. Other scattergrams such as that shown in Figures 31 and 32, for the next most effective single feature and the nominal raster period, gave little evidence of correlation. The results of these calculations and observations for other feature pairs were sufficiently encouraging that single-feature distance ratios were used to choose the next four features to be discarded from the set of eleven.

TABLE VII
Ranked Features

<u>#</u>	<u>Ratio</u>	<u>Identification</u>
902	11.68	Signal Resolution
901	10.59	Period
927	2.10	18 Hz Fourier Component
903	1.83	0 Order Poly Mean Square Fit, Bias 1
926	1.81	16 Hz Fourier Component
931	1.68	26 Hz Fourier Component
918	1.68	0 Hz Fourier Component
928	1.51	20 Hz Fourier Component
925	1.45	14 Hz Fourier Component
936	1.44	36 Hz Fourier Component
930	1.44	24 Hz Fourier Component

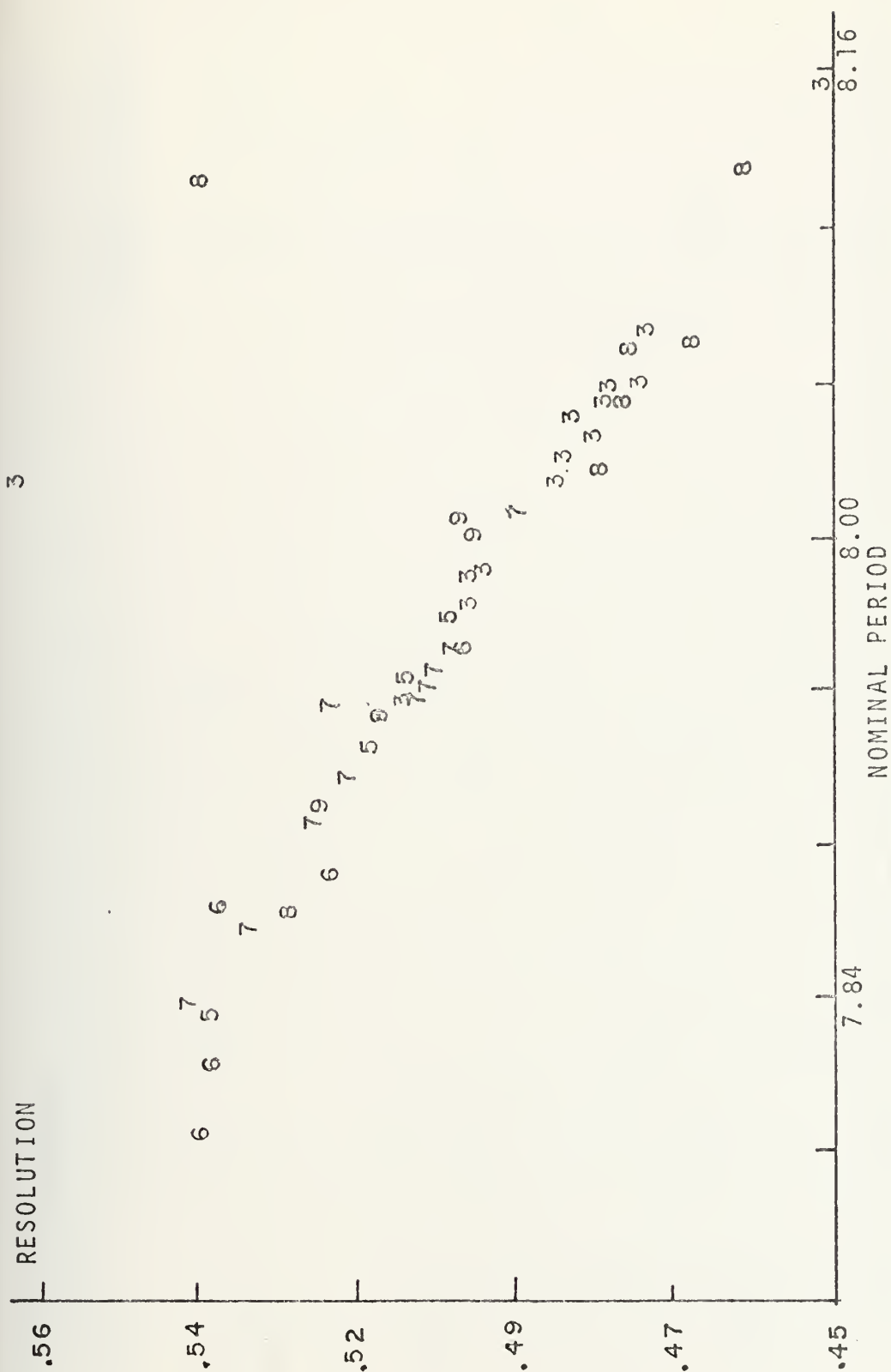
RESOLUTION



expanded in Figure 30

7.0 7.5 8.0
NOMINAL PERIOD

FIGURE 29. Scattergram of Class Membership, Nominal Period vs. Resolution



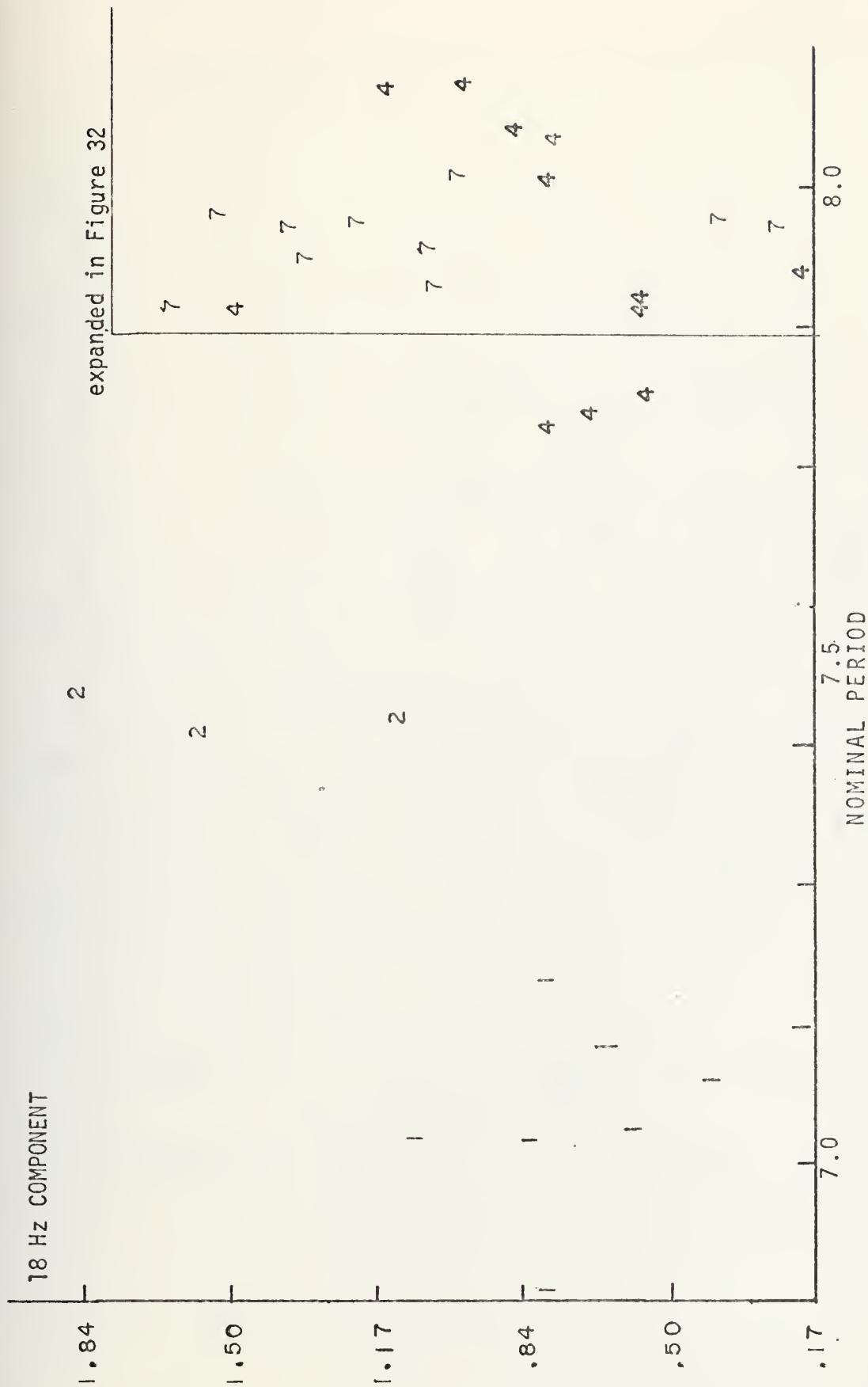


FIGURE 31. Scattergram 18 Hz Component vs. Nominal Period

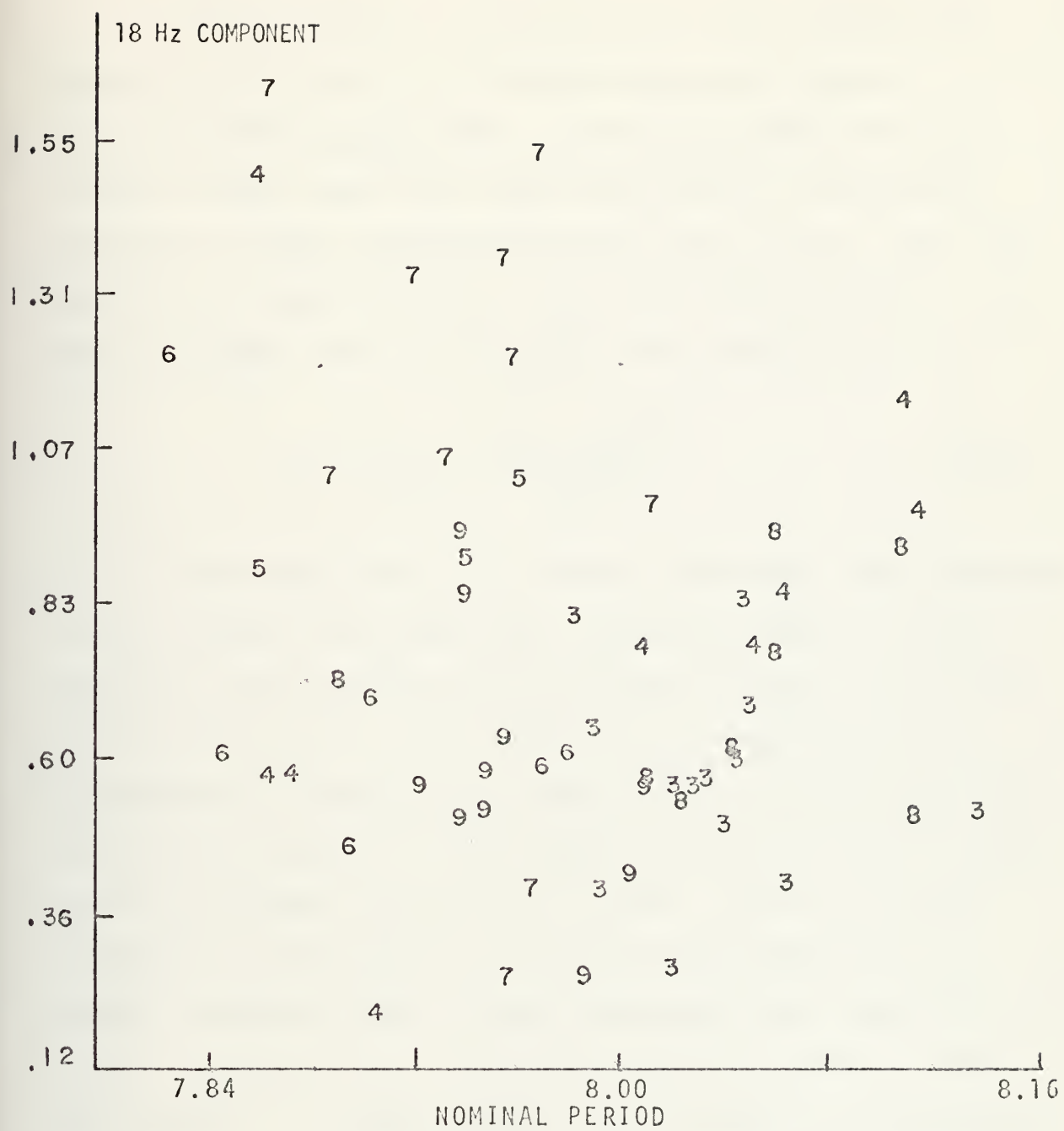


FIGURE 32. Expanded View of Figure 31

The six features finally chosen were selected by repeated application of a minimum-normalized distance classification program CNFUS. Sixty-two signals from all classes except class four were classified by the program according to their distance from the class means. This distance d_i of a particular set of n signal features, $(x_1 \dots x_n)$ from the class mean $(\mu_1^{(i)} \dots \mu_n^{(i)})$ was given by

$$d_i^2 = \frac{1}{n} \sum_{j=1}^n \frac{(x_j - \mu_j^{(i)})^2}{(\sigma_j^{(i)})^2} \quad \text{V-30}$$

The signals used for feature selection, their distance to the nearest four class means, and the confusion matrix resulting from the choice of the nearest class are shown in Tables VIII and IX. This choice of six features leads to approximately 56 percent agreement between ground truth data and the first choice of the classifier when the entire training set, including class four, is used. This, of course, represents an optimistic estimate of the probability of correct classification. Further testing of the classifier over an independent set of signal data is advisable. That the addition of features other than the basic clock rate does improve classifier performance is evident in the improvement of the 34 percent classification probability obtained using only this feature as the basis for classification. The resultant single feature confusion matrix is shown in Table X.

TABLE VIII

Classifier Selections of 4 Classes Nearest
to Signals of the Training Set

Signal	Ground Truth Class		Parameters Used							
			902	903	927	926	931	918	Classes and Distance to Mean	
427	2	2	1.16	1	1.25	4	2.52	5	4.43	
758	2	2	1.15	1	2.80					
676	2	1	.70	2	1.12	4	1.96			
770	3	3	.57	8	.73	4	.79	6	1.36	
183	3	4	.51	1	.88	8	.94	3	1.07	
192	3	1	.55	3	.84	4	.91	8	1.08	
1483	3	3	.65	4	.66	1	.84	8	.85	
1488	3	8	.74	4	.90	3	.99	1	1.23	
693	3	4	1.07	6	1.23	8	1.44	1	1.52	
1354	3	4	.72	8	.77	3	.78	9	1.08	
1367	3	3	1.01	4	1.12	1	1.15	8	1.16	
1498	3	4	1.24	3	1.25	6	1.46	8	1.49	
818	3	3	.83	4	.95	8	1.23	1	1.28	
306	3	4	.40	3	.62	8	.64	6	.82	
307	3	3	.76	4	.86	9	.98	8	1.13	
319	3	3	.96	4	1.14	8	1.32	1	1.78	
1525	3	8	.57	1	.68	4	.81	3	1.08	
930	3	3	.54	8	.56	4	.84	1	1.05	
127	5	5	.93	1	1.82	4	1.88	8	2.33	
155	5	5	.83	4	1.40	7	1.51	6	1.71	
766	5	5	.91	6	2.41	4	2.53	8	2.57	
826	5	8	.42	1	.57	4	.78	9	.89	
489	6	1	.84	4	.89	6	.92	8	1.17	
235	6	6	.88	4	.99	8	1.61	3	1.65	
236	6	6	.96	4	1.67	1	2.23	5	2.29	

787	6	6	1.21	1	1.33	4	1.56		
1543	6	8	.70	6	.73	4	.85	1	.89
1286	7	7	.80	4	1.03	1	1.20	6	1.55
1289	7	7	.88	4	1.24	1	1.41	5	1.59
672	7	7	.88	4	.94	1	1.00	9	1.30
475	7	7	.61	4	1.38	1	1.59	6	1.76
674	7	7	1.25	4	1.34	6	1.51	9	1.62
675	7	7	.95	1	.96	4	1.57	5	1.05
120	7	7	1.37	4	1.59	1	1.91	6	2.28
1384	7	1	.47	4	.79	7	.83	6	1.13
1305	7	1	.51	7	.65	4	1.23	6	1.92
141	7	1	.40	7	.99	4	1.14	5	1.53
121	9	4	.73	9	.89	7	1.05	8	1.14
754	9	1	.50	4	.53	7	.72	5	.94
145	9	1	.65	7	.82	4	.93	9	1.02
149	9	3	.75	4	.85	8	.87	9	1.02
196	9	9	.32	3	.59	4	.71	1	.74
189	9	9	.63	4	.87	3	1.08	7	1.09
201	9	8	.80	1	.80	4	.83	9	1.05
202	9	4	.56	3	.58	8	.58	6	.83
203	9	9	1.08	4	1.26	1	1.41	3	1.43
204	9	1	.93	9	1.00	4	1.25	3	1.32
1447	1	1	1.52	4	5.19	2	5.24		
1300	1	1	1.52						
875	1	1	.73	4	2.79	2	3.41	8	4.68
126	1	1	.40	4	3.75				
171	1	1	.60	4	3.22	2	4.19		
172	1	1	.79	4	3.29	2	3.41	8	5.64
775	1	1	.87	4	3.09	2	4.12		
264	1	1	.62	4	2.56	2	2.67	8	4.37
888	8	8	.61	1	.83	4	1.30	6	1.79
239	8	8	.55	3	.69	1	.84	4	.89
1487	8	8	.76	1	1.16	4	1.24	6	1.58
698	8	6	.87	4	1.05	8	1.05	1	1.18
915	8	8	.90	5	.92	1	1.24	4	1.34
446	8	1	.66	8	1.04	4	1.12	5	1.29
970	8	4	1.00	3	1.02	8	1.32	1	1.34

TABLE IX

CONFUSION MATRIX FOR CLASSIFIER USING
PARAMETERS 902, 903, 927, 926, 931, 918

		CLASSIFIER CHOICE									# missed Type I Error
		1	2	3	4	5	6	7	8	9	
Actual Class	1	8	-	-	-	-	-	-	-	-	0
	2	1	2	-	-	-	-	-	-	-	1
	3	1	-	7	5	-	-	-	2	-	8
	4	-	-	-	-	-	-	-	-	-	0
	5	-	-	-	-	3	-	-	1	-	1
	6	1	-	-	-	-	3	-	1	-	2
	7	3	-	-	-	-	-	7	-	-	3
	8	1	-	-	1	-	1	-	4	-	3
	9	3	-	1	2	-	-	-	1	3	7
# of excess Type II Error	10	0	1	8	0	1	0	5	0		

GOOD 37 MISSED 25

59% correct classification over training set.

TABLE X
CONFUSION MATRIX FOR CLASSIFIER USING CLOCK RATE ONLY

		CLASSIFIER CHOICE									# missed Type I Error
Actual Class		1	2	3	4	5	6	7	8	9	
	1	8	-	-	-	-	-	-	-	-	0
	2	-	3	-	-	-	-	-	-	-	0
	3	1	-	5	3	-	-	-	5	1	10
	4	-	-	-	-	-	-	-	-	-	0
	5	-	-	-	1	-	1	1	-	-	3
	6	1	-	-	3	-	1	-	-	-	4
	7	-	-	1	2	1	2	-	-	4	10
	8	2	-	1	-	-	1	-	3	-	4
	9	-	-	1	2	2	-	3	1	1	9
# of excess Type II Error	5	0	3	11	3	4	4	6	5		

GOOD 21 # MISSED 41

33.87% correct classification over training set.

The sequence by which the final six features were obtained involved first the calculation of the classifier confusion matrix as each of the four lowest ranked features were removed from the set of eleven. Their removal improved the classifier's performance from 50 percent to 58 percent. From this set of seven features, the effect of removing one parameter and using the remaining six yielded the following classifier performance:

<u>Parameter Removed</u>	<u>Probability of Correct Classification</u>
None	58 %
918	25 %
931	53 %
926	50 %
903	51 %
927	53 %
901	59 %
902	56 %

The above suggests that the best single feature to be removed at this stage is 901, the nominal raster period. That this remaining group of six features is not necessarily the optimal choice, can be shown by observing that a choice of five features, including parameter 901 leads to a 61 percent probability of correct identification shown in Table XI.

TABLE XI

CONFUSION MATRIX FOR CLASSIFIER
PARAMETERS USED 901, 903, 926, 931, 918

		CLASSIFIER CHOICE									# missed Type I Error
		1	2	3	4	5	6	7	8	9	
Actual Class	1	8	-	-	-	-	-	-	-	-	0
	2	1	2	-	-	-	-	-	-	-	1
	3	1	-	6	4	-	-	-	4	-	9
	4	-	-	-	-	-	-	-	-	-	0
	5	-	-	-	-	3	-	-	1	-	1
	6	1	-	-	-	-	4	-	-	-	1
	7	2	-	-	1	-	-	7	-	-	3
	8	1	-	-	1	1	-	-	4	-	3
	9	2	-	-	3	-	-	-	1	4	6
# Excess Type II Error	8	0	0	9	1	0	0	6	0	0	

GOOD 38 # MISSED 24

61.29% CORRECT CLASSIFICATION OVER TRAINING SET

VI. SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

It has not been the intent of this thesis to suggest feature measurements which can replace those currently used in the Parameter Encoder's source identification scheme. The fact that the measurement process discussed in this thesis is currently unable to provide meaningful measurements for about 20% of the signal data attempted is but one reason for not adopting the indicated Fourier components and polynomial coefficients as standard features to be measured.

These features however, constitute a hitherto unexploited source of raster pattern information. The use of four of these features in conjunction with the nominal raster period or signal clock period provides approximately 100% improvement in classifier capability over the separation performance of the nominal raster period alone. In view of the amount of overlap in the feature distributions for several classes, reflected in both the scattergrams and the distances from the individual signals to the nearest class means, it is not advisable to use more than four or five such features. Neither is it advisable to draw too firm a conclusion as to the optimal choice of these new features. The 62 test and training signals do not by any stretch of the imagination constitute a completely representative statistical base.

It is rather surprising that those terms intended to characterize the mean bias, variance measures and transient

phenomena show as little single-feature separating capability as they do, since it was precisely these phenomena which have been used by system operators as visual aids to signal identification. While one might reasonably associate the zero order polynomial mean square fit coefficients and the zero-Hz Fourier component with transient related effects, the higher-order Fourier series terms have not previously been noted as particularly prominent in the raster signal display. In view of the series of nonlinearities present in obtaining the interpolated minimal phase raster points, it is not impossible that these higher order terms owe their separating capability to the data modulation of the signal source. In this respect the "sideband" structure of the single feature separation data of Figure 27 is most intriguing.

Before completely discounting the usefulness of bias measures and transient phenomena, a more careful representation and analysis of the data in the transient region should be attempted. The representation could perhaps involve the calculation of the minimal phase by use of a moving-average technique which works backward from data in the steady-state region to the starting time of the signal. At the same time use of a different decay constant and a smoother integration to obtain the Laguerre polynomial coefficients might also prove of value.

If additional accurate ground truth data is available, the modeling of the probability distributions by something

other than a Gaussian form could be useful. The average squared magnitude of the Fourier components might be more accurately represented by a Rayleigh distribution. Although some of the measured parameters for one or more classes may not have a unimodal distribution, the number of accurately labeled samples currently available provides little encouragement for the use of Parzen window probability estimators.

The foregoing comments suggest that four or five parameters characterizing the raster scan data may provide signal identifying information useful to about the same degree of accuracy as visual interpretation of the raster scan display. Since the optimal parameters measured in this investigation do not appear to correspond to the visual cues used by system operators, one has reason to believe that other parameters derived in subsequent analysis might improve upon this degree of accuracy.

The measurement technique investigated exists outside the mainstream of the Parameter Encoder's identification process. As such it constitutes a relatively slow procedure (one minute per signal), requiring the use of disc data files which contain much of the same information available directly from the master file structure. Even should the programs used be modified to work solely with the master file, the data obtained might be of greatest use in a sequential identification scheme. It could be used much as the present raster scan display is; to resolve ambiguities in the choice between a relatively small number of classes.

APPENDIX A

METHODS OF PROBABILITY ESTIMATION

1. Parametric Techniques

Perhaps the most common technique which has been used to infer the a priori probability structure presupposes that the distribution of features in a given class is multivariate normal. As a first step the mean and variance for each feature of a class are estimated. For the class C_i having samples $(\underline{x}_1^i \dots \underline{x}_{K_i}^i)$, these estimates may be:

$$\underline{\mu}^{(i)} = \frac{1}{K_i} \sum_{j=1}^{K_i} \underline{x}_j^{(i)} \quad \text{A-1}$$

$$\sigma_k^{(i)2} = \frac{1}{K_i - 1} \sum_{j=1}^{K_i} (x_{jk} - \mu_k^{(i)})^2 \quad \text{A-2}$$

where $\mu_k^{(i)}$ and $\sigma_k^{(i)2}$ are components of $\underline{\mu}^i$ and $\underline{\sigma}^{2(i)}$. The variance vector represents a simplification from the more general multivariate form:

$$p(\underline{x}|C_i) = \frac{1}{2\pi^{d/2} |\underline{S}_i|} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}^i)^T \underline{S}_i^{-1} (\underline{x} - \underline{\mu}^i)\right] \quad \text{A-3}$$

The components of the vector $\underline{\sigma}^{2(i)}$ are the diagonal elements of the $d \times d$ covariance matrix, which may be estimated by:

$$\underline{S}_i = \frac{1}{K_i - 1} \sum_{j=1}^{K_i} (\underline{x}_j^{(i)} - \underline{\mu}_j^{(i)}) (\underline{x}_j^{(i)} - \underline{\mu}_j^{(i)})^T \quad \text{A-4}$$

Under the assumption of independent features, the distribution is characterized by the product of d Gaussian distributions.

$$p(\underline{x}|C_i) = \prod_{k=1}^d \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_k^{(i)}} \exp\left[-\frac{1}{2} \frac{1}{(\sigma_k^{(i)})^2} (x_k - \mu_k^{(i)})^2\right] \quad A-5$$

If required, the technique may be extended to include the estimation of a covariance matrix for each class by use of Equations A-3 and A-4. This procedure has enjoyed great popularity, particularly in the modeling of features with unimodal class conditional probability distributions. The ease with which the probability measure of a Gaussian distribution may be interpreted as a weighted distance measure has been an important factor in its retention.

2. Non-Parametric Techniques

Another technique for obtaining the a-priori probability density function involves its direct estimation through the use of window functions or potential functions. In this technique, the estimate at some point \underline{x} in the d -dimensional feature space is obtained by the superposition of terms which are generalized distance functions of that point and the training set points $\underline{y}_j^{(i)}$ of the class C_i .

$$p(\underline{x}|C_i) = \frac{1}{K_i} \sum_{j=1}^{K_i} \gamma(\underline{x}, \underline{y}_j^{(i)}) \quad A-6$$

The technique arises from a viewpoint proposed by Parzen. In one-dimensional cases, one can estimate the probability density at some point x by counting the number of samples $N_h(x)$ in the interval $[x-h, x+h]$ and dividing by the product of the total number of samples M and the interval width $2h$

$$\hat{p}_M(x) = \frac{N_h(x)}{M2h} \quad \text{A-7}$$

If one defines a "window" function

$$K(y) = \begin{cases} \frac{1}{2}, & |y| \leq 1 \\ 0 & |y| > 1 \end{cases} \quad \text{A-8}$$

one can write

$$\hat{p}_M(x) = \frac{1}{Mh} \sum_{j=1}^M K\left(\frac{x-y_j}{h}\right) \quad \text{A-9}$$

This corresponds to the potential function

$$\gamma(x, y) = \frac{1}{h} K\left(\frac{x-y}{h}\right) \quad \text{A-10}$$

Additionally, the argument presented may be extended to multiple dimensions by considering

$$K(\underline{u}) = \begin{cases} 1, & |u_j| \leq \frac{1}{2} \text{ for } j = 1, \dots, m \\ 0, & \text{otherwise} \end{cases}$$

$$(\underline{x}, \underline{y}) = \frac{1}{V_m} K\left(\frac{\underline{x} - \underline{y}}{h_m}\right) \quad A(11-12)$$

where V_m is a multidimensional volume of the hypercube of side h_m .

The window function K need not be the exact form of Equation A-11. It can be shown that if the estimate $\hat{p}_M(\underline{x})$ is a function both of the total number of samples M and $V_m = V_m(M)$, the conditions

$$0 \leq K(\underline{y}) < \infty$$

$$\int K(\underline{y}) d\underline{y} = 1$$

A(13-16)

$$\lim_{M \rightarrow \infty} V_m = 0$$

$$\lim_{M \rightarrow \infty} M V_m = \infty$$

ensure that as the number of total samples $M \rightarrow \infty$

$$\lim_{M \rightarrow \infty} \hat{p}_M(\underline{x}) = p(\underline{x}) \quad A-17$$

$$\lim_{M \rightarrow \infty} (\hat{p}_M(\underline{x}) - p(\underline{x}))^2 = 0 \quad A-18$$

Thus, we can consider $\hat{p}_M(\underline{x})$ to be a "blurred" or "noisy" version of $p(\underline{x})$ as seen through an averaging window.

Intuitively, one can describe the following useful properties of the potential function: $\gamma(\underline{x}, \underline{y})$

1. $\gamma(\underline{x}, \underline{y})$ should be maximum for $\underline{x} = \underline{y}$.
2. $\gamma(\underline{x}, \underline{y})$ should be approximately zero for \underline{x} distant from \underline{y} .
3. $\gamma(\underline{x}, \underline{y})$ should be continuous and decrease monotonically with the distance between \underline{y} and \underline{x} .
4. If $\gamma(\underline{x}_1, \underline{y}) = \gamma(\underline{x}_2, \underline{y})$ where \underline{y} is a sample point, the patterns represented by \underline{x}_1 and \underline{x}_2 should have about the same similarity to \underline{y} .

Such properties can be obtained with a potential function of the form:

$$\gamma(\underline{x}, \underline{y}) = \frac{1}{2\pi^{d/2} \sigma_1 \dots \sigma_d} \exp\left[-\frac{1}{2} \sum_{i=1}^d \frac{(x_i - y_i)^2}{\sigma_i^2}\right] \quad A-19$$

where σ_i are arbitrarily chosen values. Although the form here is Gaussian, and the function of a distance measure, it is also possible to construct a potential function of the form:

$$\gamma(\underline{x}, \underline{y}) = \sum_{i=1}^N \lambda_i^2 \phi_i(\underline{x}) \phi_i(\underline{y}) \quad A-20$$

where $\{\phi_i\}$ is a complete set of multivariate orthonormal

functions and λ_i are constants. Such techniques have been described more completely in the work of Aizermann and Bravermann and others [1, 7, 19].

Given sufficient samples, the Parzen window or potential function approach essentially assures satisfactory convergence to an arbitrarily complex distribution. Unfortunately, this sufficient number of samples may be far greater than the number required if the form of the distributions was known. Since every sample point is used in the construction of the density function, the above approach affords little economy in the way of data reduction and leads to a demand for computation time and storage space exponentially increasing with the number of features.

APPENDIX B

CLASSIFICATION ALGORITHM

1. Distance Based Classification Schemes

An assumption commonly used to implement a decision rule is that of multivariate normal class dependent density functions of the form:

$$\begin{aligned} p(\underline{x}|C_i) &= p(C_i) \frac{1}{2\pi^{d/2} |\underline{S}_i|} \exp \left[-\frac{1}{2} [\underline{x} - \underline{\mu}^{(i)}]^T \underline{S}_i^{-1} [\underline{x} - \underline{\mu}^{(i)}] \right] \\ &= p(C_i) \frac{1}{V_i} \exp \left[-\frac{1}{2} [\underline{x} - \underline{\mu}^{(i)}]^T \underline{S}_i^{-1} [\underline{x} - \underline{\mu}^{(i)}] \right] \quad B-1 \end{aligned}$$

where

\underline{x} is the d-component feature vector

$\underline{\mu}^i$ is the d-component class mean vector

\underline{S}_i is a d x d class covariance matrix

$|\underline{S}_i|$ is the determinant of \underline{S}_i .

For such a distribution one may obtain the locus of points of constant density as a hyperellipsoid for which the quadratic surface $(\underline{x} - \underline{\mu}^{(i)})^T \underline{S}_i^{-1} (\underline{x} - \underline{\mu}^{(i)})$ is constant.

The quantity

$$r_i^2 = (\underline{x} - \underline{\mu}^{(i)})^T \underline{S}_i^{-1} (\underline{x} - \underline{\mu}^{(i)}) \quad B-2$$

is sometimes called the Mahalanobis distance from \underline{x} to $\underline{\mu}^{(i)}$.

Since the probability densities are always non-negative, the maximum likelihood criteria

$$p(C_i)p(C_i|\underline{x}) \geq p(C_i)p(C_j|\underline{x}) \quad \text{for all } j \quad \text{B-3}$$

implies

$$\log p(C_i)p(C_i|\underline{x}) \geq \log p(C_i)p(C_j|\underline{x}) \quad \text{B-4}$$

Substituting the multivariate density from Equation 1, the above expression yields

$$\begin{aligned} & \log p(C_i) + \log V_i - \frac{1}{2}[\underline{x} - \underline{\mu}^i]^T \underline{S}^{-1} [\underline{x} - \underline{\mu}^i] \\ & \geq \max_{j \neq i} \{ \log p(C_j) - \log V_j - \frac{1}{2}[\underline{x} - \underline{\mu}^{(j)}]^T \underline{S}_j^{-1} [\underline{x} - \underline{\mu}^{(j)}] \} \end{aligned} \quad \text{B-5}$$

The above equation defines the decision boundaries of each class. If a common covariance matrix $\underline{S}_i = \underline{S}_j = \underline{S}$ is assumed for each class in the above equation, the $\log V_i$ terms may be ignored and the form of the decision boundary becomes

$$\begin{aligned} & \frac{1}{2}(\underline{x} - \underline{\mu}^{(i)})^T \underline{S}^{-1} (\underline{x} - \underline{\mu}^{(i)}) \\ & \leq \min_{j \neq i} \{ (\underline{x} - \underline{\mu}^{(j)})^T \underline{S}^{-1} (\underline{x} - \underline{\mu}^{(j)}) \} + \log \frac{p(C_i)}{p(C_j)} \end{aligned} \quad \text{B-6}$$

Similarly, if equiprobable, a-priori class probabilities are assumed, $p(C_i) = p(C_j)$ then the decision rule becomes

$$11) \quad [\underline{x} - \underline{\mu}^{(i)}]^T \underline{S}^{-1} [\underline{x} - \underline{\mu}^{(i)}] = \min_{j \neq i} \{ [\underline{x} - \underline{\mu}^{(j)}]^T \underline{S}^{-1} [\underline{x} - \underline{\mu}^{(j)}] \} \quad B-7$$

This is equivalent to simply assigning the feature vector \underline{x} to that class whose pattern, $\underline{\mu}^{(i)}$, is at the minimum Mahalanobis distance from \underline{x} . The concept of Mahalanobis distance is particularly applicable to well-separated unimodal feature distributions, but presents a considerable problem in those cases where the distributions exhibit more than one local maximum.

2. K - Nearest Neighbor Algorithms

The concept of distance may be applied in a slightly different form to probability functions not well characterized by the multivariate normal form. In this application, distances in feature space are quite frequently normalized by some form of global variance or maximal spread. The normalized Euclidean distance from the sample point \underline{x} to each point of the training set $(\underline{x}_1^{(1)} \dots \underline{x}_{K_1}^{(1)} \dots, \underline{x}_1^{(m)} \dots \underline{x}_{K_m}^{(m)})$ is calculated and ranked in order of increasing distance. In the simplest algorithm, class membership of the sample point is determined by the class membership of the greatest number of the k-nearest neighbors in the training set. For example, suppose that of the ten points of the training set nearest

to the test sample, three were members in class 1, four of class 2 and one each were members of classes 4, 5, 6, the simplest algorithm would choose class 2 as the probable classification. Alternative algorithms of this type weight the contribution of the k-nearest training points according to their distance ranking. The nearest point might be assigned a weighted class vote of 10, the next nearest, a weight of 9 and so forth. Class membership of the sample point would then be determined by summing the weighted class votes of the ten nearest neighbors and choosing that class which received the maximum number of votes.

Such procedures are reminiscent of the Parzen window estimators described in Appendix A.2, and in fact can be shown to be equivalent if, instead of rank order, distance weighted contributions to the class vote are allowed. Numerous techniques [3, 4, 10] exist to increase the computational efficiency of these K-nearest neighbor algorithms by reducing the number of Euclidean distance terms which must be calculated for each sample point. It must be mentioned that these techniques are of greatest use when the individual classes are equiprobable or at the very least, when every class of the training set has more than k elements.

APPENDIX C

FEATURE SELECTION AND RANKING

1. Linear Combinations of Features

The objectives of low dimensionality and retention of sufficient information may often be obtained through the use of custom orthonormal transformations of the measurement space [13, 15, 19, 20]. In such an interpretation the sample data is presumed to be adequately represented by the linear combination of some finite set of functions $\phi_1 \dots \phi_n$

$$\underline{z} = \sum_{i=1}^n a_i \underline{\phi}_i \quad \text{C-1}$$

where

$$\underline{\phi}_i = (\phi_{i1}, \dots, \phi_{im})$$

For a given set of functions the weights are adjusted so that the average mean square error

$$\sum_{k=1}^n (z_n - \sum_{i=1}^n a_i \phi_{ik})^2 \quad \text{C-2}$$

is minimized.

The values a_i are thus determined by taking partial derivatives with respect to a_i and setting the result equal to zero, obtaining n equations in n unknowns of the form

$$\sum_{i=1}^n a_i \sum_{k=1}^m \phi_{ik} \phi_{jk} = \sum_{k=1}^m z_k \phi_{ik} \quad \text{C-3}$$

Now if the functions $\phi_1 \dots \phi_n$ are orthonormal; that is if

$$\sum_{k=1}^m \phi_{ik} \phi_{jk} = \begin{matrix} 0 & i \neq j \\ 1 & i = j \end{matrix} \quad \text{C-4}$$

Equation C-3 reduces to

$$a_i = \sum_{k=1}^m z_k \phi_{ik} \quad \text{C-5}$$

Then values of a_i may then be used as the pattern space \underline{x} representation of the vector z under the linear transformation

$$\underline{x} = \underline{T} \underline{z} \quad \text{C-6}$$

$$\underline{T} = [\phi_1, \phi_2, \dots, \phi_n]$$

Expansion of a function by a Fourier series represents one form of orthonormal transformation, but one which has distinct advantages in computational efficiency.

It can be shown that Karhunen-Louve expansion of N m -dimensional samples $z_i = (z_{i1}, \dots, z_{im})$ leads to a $m \times m$ eigenvalue problem.

$$K \phi_k = \sigma_k^2 \phi_k \quad \text{C-7}$$

where

$$k_{rs} = \frac{1}{N} \sum_{i=1}^N z_{ir} z_{is} \quad \text{C-8}$$

$$\underline{K} = [k_{rs}]$$

The above solution does in fact minimize mean square error averaged over the entire sample set. By choosing eigenvectors corresponding to the n largest eigenvalues, one can achieve dimensionality reduction from m to n . This approach, also known as principal factor analysis, requires the computation of the eigenvectors and eigenvalues of a large ($m \times m$) matrix. Note that if the number of signals N is less than m , there will be at most $(N-1)$ non-zero eigenvalues, and the minimum mean square error is zero.

A simpler procedure is to choose a small number of linearly independent patterns typical of the different classes or types of samples; then apply the Gram-Schmidt orthonormalization process to them directly. In this process one begins with the "typical" samples, perhaps derived from class means.

$$\begin{aligned} \underline{\mu}^{(1)} &= (\mu_1^{(1)} \dots \mu_m^{(1)}) \\ \underline{\mu}^{(2)} &= (\mu_1^{(2)} \dots \mu_m^{(2)}) \\ &\vdots \\ \underline{\mu}^{(n)} &= (\mu_1^{(n)} \dots \mu_m^{(n)}) \end{aligned} \quad \text{C-9}$$

and constructs the orthonormal functions ϕ_i in the following manner:

$$\begin{aligned}
 \psi_1 &= \underline{\mu}^{(1)} \\
 \phi_1 &= (\psi_1, \psi_1)^{-\frac{1}{2}} \psi_1 \\
 c_{12} &= (\phi_1, \underline{\mu}^{(2)}) \\
 \psi_2 &= \underline{\mu}^{(2)} - c_{12} \phi_1 \\
 &\vdots \\
 c_{jk} &= (\phi_j, \underline{\mu}^{(k)}) \\
 \psi_k &= \underline{\mu}^{(k)} - c_{1k} \phi_1 - c_{2k} \phi_2 \cdots - c_{k-1,k} \phi_{k-1} \\
 \phi_k &= (\psi_k, \psi_k)^{-\frac{1}{2}} \psi_k
 \end{aligned} \tag{C-10}$$

where

$$(\underline{g}_j, \underline{g}_k) = \frac{1}{m} \sum_{i=1}^m g_{ji} g_{ki} \tag{C-11}$$

Clearly, all of the samples used to generate the orthonormal vectors may be represented exactly by an n-term expansion; thus the representation error of the 'typical' sample is minimized. If these samples are truly typical, one can

expect other samples to be represented closely by the orthonormal basis functions.

The dimensionality of the resultant pattern space is the same as the number of linearly independent typical samples.

2. Criteria for Feature Selection and Ranking

Rather than achieving a pattern space through use of a transformation which minimizes mean square error over the training set, it is often desirable to optimize some measure of separation quality. Meisel [14] catalogs more than ten different measures of separation quality which are directly calculable from general a priori conditional density functions. For unimodal distributions, however, it is convenient to describe such a measure in terms of average distances in the feature space.

Two distance measures need to be considered. One, the inter-class distance, might be characterized by either the distance between the means of the classes or the average squared distances between points of different classes. Another, the intra-class distance, is the measure of the internal scatter of samples and may be characterized by an average squared distance between the points of each class and summed over the classes. In the two-class problem, a common definition of the interclass distance, S , and intra-class distance, R_i , is

$$S = \frac{1}{K_1 K_2} \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} d^2[\underline{x}_i^{(1)}, \underline{x}_j^{(2)}] \quad C-12$$

$$R_i = \frac{2}{K_i K_{i-1}} \sum_{j=1}^{K_i} \sum_{k=j}^{K_i} d^2[\underline{x}_j^{(i)}, \underline{x}_k^{(i)}] \quad C-13$$

where $d^2(x,y)$ is some distance measure of the vector variables x,y .

Conventional distance measures might use one of the following forms:

$$\text{Euclidian} \quad d_1^2(\underline{x}, \underline{y}) = \sum_{i=1}^n (x_i - y_i)^2$$

$$\text{city block} \quad d_2(\underline{x}, \underline{y}) = \sum_{i=1}^n |x_i - y_i|$$

$$\text{Mahalanobis} \quad d_3^2(\underline{x}, \underline{y}) = (\underline{x} - \underline{y})^T \tilde{A}^{-1} (\underline{x} - \underline{y})$$

$$\text{localized distance} \quad d_4^*(x,y) = 1 - \exp[-\frac{1}{2D} d_1^2(\underline{x}, \underline{y})] \quad C(14-17)$$

In generalizing this distance criteria to a m-class problem, it is convenient to resort to the concept of scatter matrices where

$$\tilde{S}_i = \sum_{k=1}^{K_i} (\underline{x}_k - \underline{\mu}^i) (\underline{x}_k - \underline{\mu}^i)^T \quad C-18$$

$$\tilde{S}_w = \sum_{i=1}^m \tilde{S}_i \quad C-19$$

represent the individual class scatter matrices and the collective within-class scatter matrices respectively. If one defines a total scatter matrix S_T in terms of a global mean \underline{M}

$$\underline{M} = \frac{1}{\sum_i K_i} \sum_{i=1}^m K_i \underline{\mu}^{(i)} \quad \text{C-20}$$

$$S_T = \sum_{\text{all samples}} (\underline{x} - \underline{M})(\underline{x} - \underline{M})^T \quad \text{C-21}$$

we find that

$$\begin{aligned} S_T &= \sum_{i=1}^m \sum_{\underline{x} \in C_i} (\underline{x} - \underline{\mu}^{(i)})(\underline{x} - \underline{\mu}^{(i)})^T + \sum_{i=1}^m \sum_{\underline{x} \in C_i} (\underline{\mu}^{(i)} - \underline{M})(\underline{\mu}^{(i)} - \underline{M})^T \\ &= S_W + S_B \end{aligned} \quad \text{C-22}$$

provides a natural definition for both inter-set and intra-set scatter matrices:

$$S_W = \sum_{i=1}^m S_i \quad \text{C-23}$$

$$S_B = \sum_{i=1}^m K_i (\underline{\mu}^{(i)} - \underline{M})(\underline{\mu}^{(i)} - \underline{M})^T \quad \text{C-24}$$

These three matrices not only provide the basis for multiple discriminant analysis [7, 14, 18], but their trace or

determinant can also serve as a measure of separability for use in an algorithm involving either linear combinations of features or discrete feature selection. For example, as criteria for separability, one might choose to maximize the quality function

$$Q_1 = \text{tr } \tilde{S}_W^{-1} \tilde{S}_B \quad \text{C-25}$$

or to minimize one of the quality functions

$$Q_2 = \text{tr } \tilde{S}_T^{-1} \tilde{S}_W \quad \text{C-26}$$

$$Q_3 = \frac{|\tilde{S}_W|}{|\tilde{S}_T|} \quad \text{C-27}$$

All of the above criteria are particularly interesting in that they can be evaluated as a result of multiple discriminant analysis using linear combinations of features.

For a c -class problem we can accomplish the projection of the d -dimensional feature space onto some alternative space of not more than $c-1$ dimensions through the use of a non-unique $d \times (c-1)$ matrix \tilde{W} .

$$\underline{y} = \tilde{W}^T \underline{x} \quad \text{C-28}$$

It is straightforward problem to show that the resultant scatter matrices \tilde{S}_W, \tilde{S}_B in this projection space are

$$\tilde{S}_w = W^T \tilde{S}_w W$$

$$\tilde{S}_B = W^T \tilde{S}_B W \quad C-29$$

If one then chooses W to maximize the ratio of between-to within-class scatter matrix determinants,

$$Q_4 = \frac{|\tilde{S}_B|}{|\tilde{S}_w|} = \frac{|W^T \tilde{S}_B W|}{|W^T \tilde{S}_w W|} \quad C-30$$

there is a solution of the form

$$\tilde{S}_B \underline{w}_i = \lambda_i \tilde{S}_w \underline{w}_i. \quad C-31$$

If \tilde{S}_w is non-singular one can directly solve an eigenvalue problem related in form to the Karhunen-Louve problem Equation C-7.

$$(\tilde{S}_w^{-1} \tilde{S}_B) \underline{w}_i = \lambda_i \underline{w}_i \quad C-32$$

or one may then solve the characteristic polynomial

$$|\tilde{S}_B - \lambda_i \tilde{S}_w| = 0 \quad C-33$$

for the largest eigenvalues and find the corresponding eigenvectors from

$$(\underline{S}_B - \lambda_i \underline{S}_w) \underline{w}_i = 0$$

C-34

If $m-1$ of the vectors $(\underline{\mu}^{(i)} - \underline{M})$ are linearly independent, there will be not more than $m-1$ non-zero eigenvalues for this problem. If the within-class scatter matrix is isotropic, which is the case if all features are independent and suitably normalized, the resultant eigenvectors span the space defined by the vector set $(\underline{\mu}^{(i)} - \underline{M})$ and may be established by Gram-Schmidt orthogonalization. The eigenvalues of this problem serve to characterize the trace and determinant criteria. For example, the use of all features leads to the criteria values

$$Q_1 = \text{tr } \underline{S}_w^{-1} \underline{S}_B = \sum_{i=1}^{m-1} \lambda_i$$

$$Q_2 = \text{tr } \underline{S}_T \underline{S}_w = \sum_{i=1}^{m-1} \frac{1}{1 + \lambda_i}$$

$$Q_3 = \frac{|\underline{S}_w|}{|\underline{S}_T|} = \prod_{i=1}^{m-1} \frac{1}{1 + \lambda_i} \quad \text{C(35-37)}$$

As was previously mentioned, the criteria described here are best suited to unimodal distributions and increase the theoretically achievable error rate. In those cases where it can be established that a unimodal distribution is inappropriate, the use of separability measures more closely related to the estimated conditional probability density is indicated [16]. For example one might use a quality measure:

$$Q_5 = \frac{1}{N} \sum_{j=1}^N [p^2(\underline{F}(\underline{x}_j)) - \sum_{i=1}^m [p(C_i) p(\underline{F}(\underline{x}_j) | C_i)]^2 \quad C-38$$

where N is the total number of training set samples and $\underline{F}(\underline{x})$ represents some selection of features from the initial feature space and

$$P(\underline{x}) = \sum_{i=1}^m p(C_i) p(\underline{x} | C_i) \quad C-39$$

Q_5 is in fact a measure of overlap and has an optimal value of zero.

3. Feature Search Algorithms

A cursory inspection of the problem of finding those n of m features which optimize one of the criteria given in the previous section shows that

$$N_t = \frac{m!}{n! (m-n)!} \quad C-40$$

evaluations must be performed if one is to conduct an exhaustive search of all possible feature combinations. The exhaustive search for the best ten or twenty features for example, requires consideration of more than 184,000 possibilities. Typically, the practical feature search employs suboptimal procedures that may be justified if simplifying assumptions regarding the nature of the feature distributions are allowed.

The simplest suboptimal search procedure is to assume the independence of features and evaluate the desired criteria function for all single features. These m features are then ranked in order of decreasing optimality and the first n are chosen. The number of criteria evaluations is the same as the dimension of initial feature space.

A second search method used by Mucciardi and Gose [17] involves a technique known as search without replacement. This method requires N_E criteria evaluations where

$$N_E = \frac{n(m-n)}{2} \quad \text{C-41}$$

Fu [9] has proposed a technique of sequential feature selection which can be modified to include the cost of feature measurement and Chang [3] has developed alternative dynamic programming approaches to feature selection, one of which requires only a slightly greater number of evaluations N_E than the search without replacement algorithm.

$$N_E = n\left(m - \frac{n-1}{2}\right) + n - 2(m-1) \quad \text{C-42)$$

The single feature ranking technique has the disadvantage of ignoring the effects of feature correlations. The search without replacement technique provides a method for treating these correlations, but it also presumes that the features obtained through n stages of conditional single

feature evaluation procedure are the n best features, which need not be the case. In general, as one progresses to successively more complicated search procedures, more of the effects of feature dependence may be taken into account.

PARAMETER HISTOGRAMS

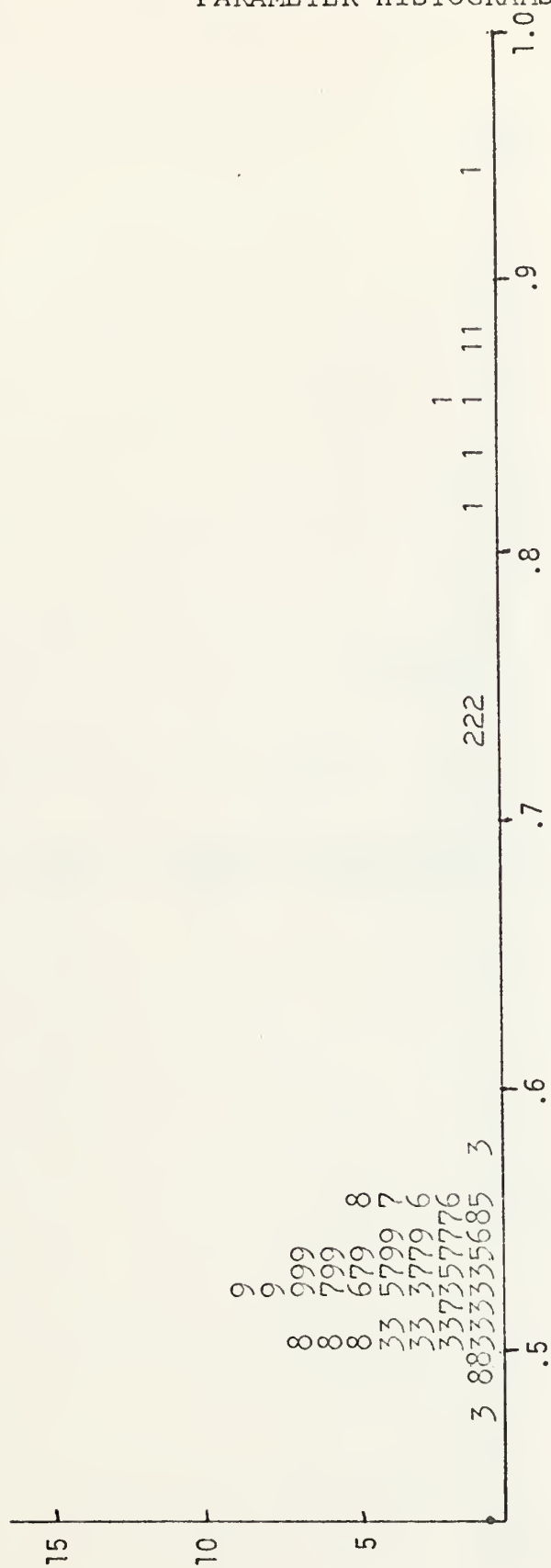


FIGURE 33. Histogram; Resolution (902)

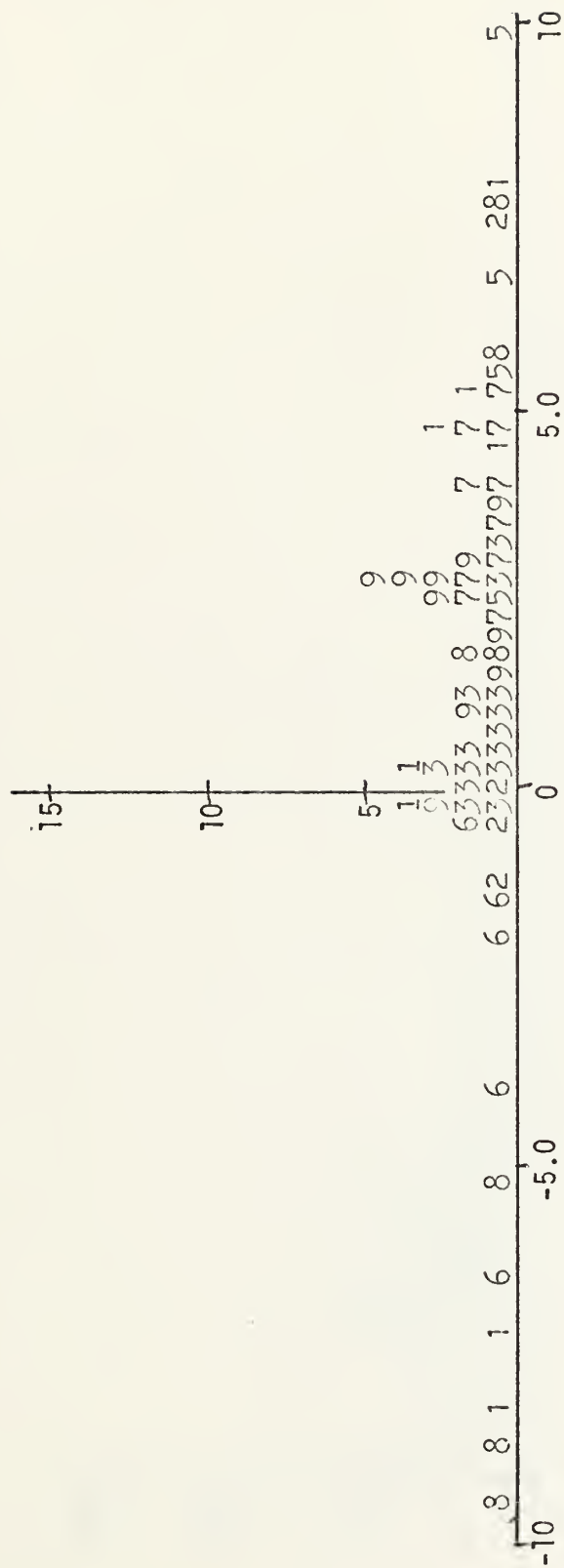


FIGURE 34. Histogram: Zero Order Coefficient of Polynomial Fit (903)

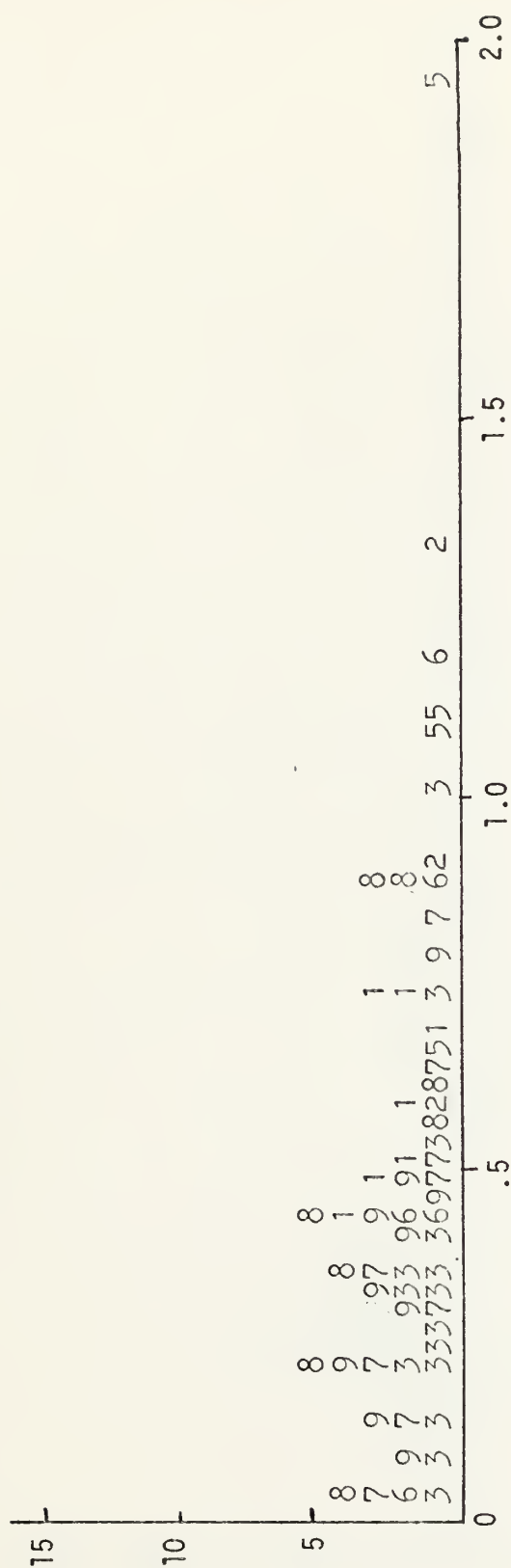


FIGURE 35. Histogram: 0 Hz Frequency Component (918)

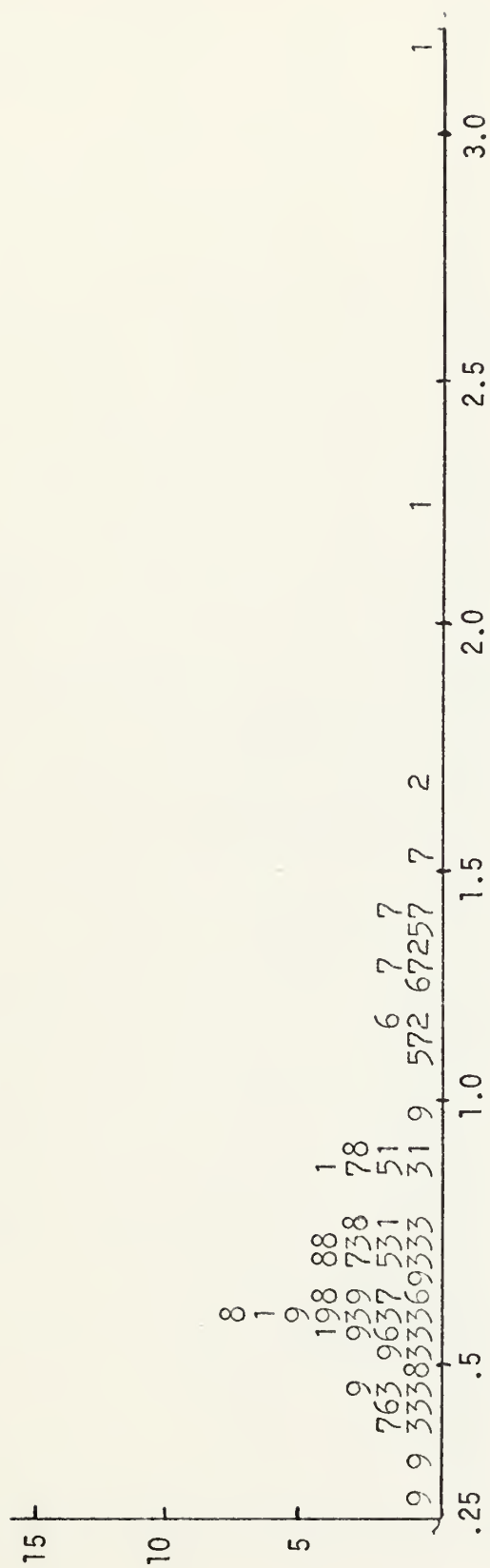


FIGURE 36. Histogram: 18 Hz Frequency Component (926)

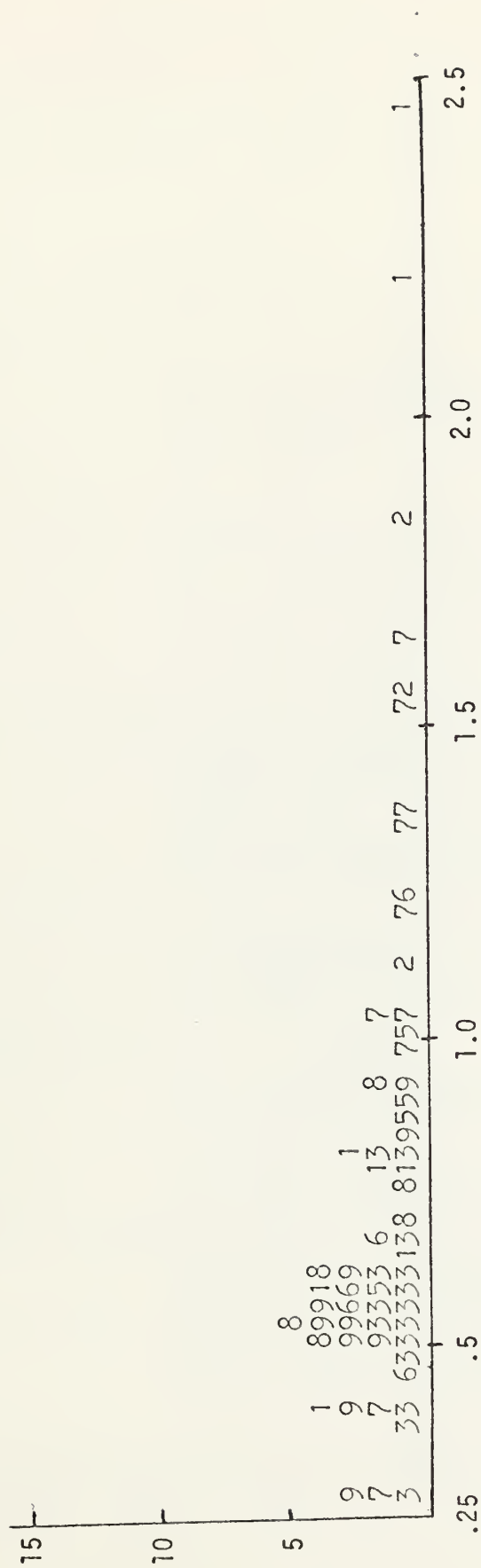


FIGURE 37. Histogram: 20 Hz Frequency Component (927)

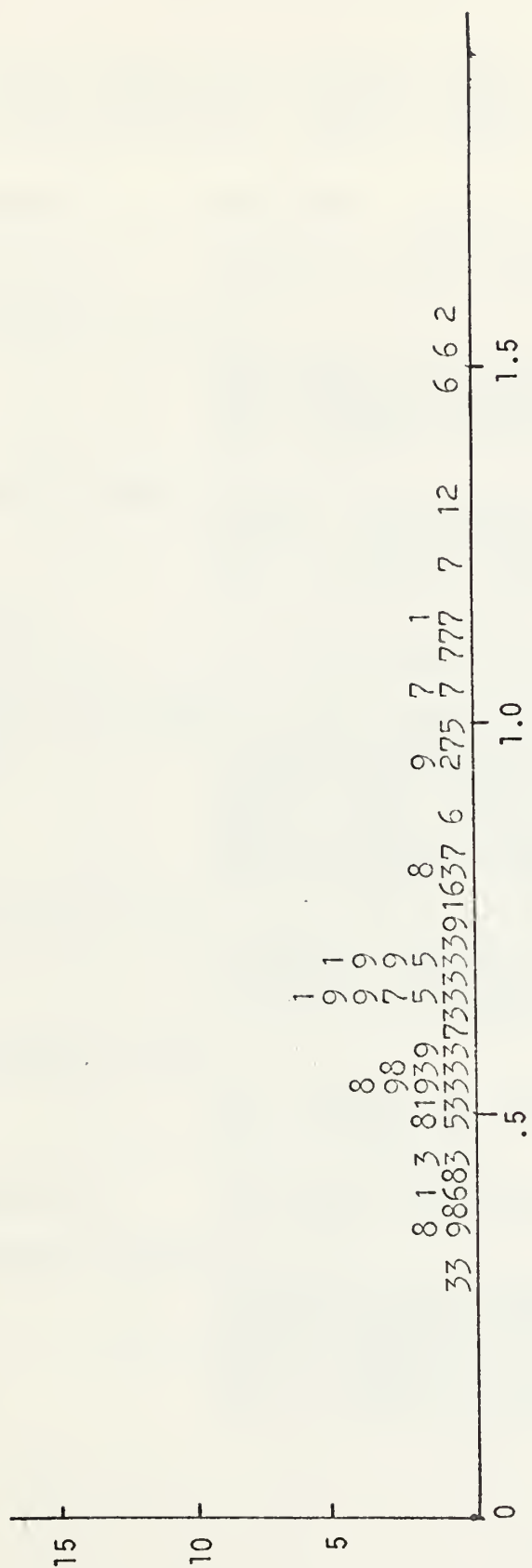


FIGURE 38. Histogram: 28 Hz Frequency Component (931)

COMPUTER PROGRAMS

THE PROGRAMS DISCUSSED IN THIS SECTION ALL USE LIBRARY AND EXECUTIVE ROUTINES WHICH ARE PART OF THE PARAMETER ENCODER SOFTWARE PACKAGE. AS AN AID TO INTERPRETATION A BRIEF DESCRIPTION OF SOME OF THESE ROUTINES FOLLOWS

ROUTINE	FUNCTION
ALPHA	CONTROLS TYPE OF OUTPUT TO CRT
CHOUT(I)	OUTPUTS SINGLE CHARACTER TO CRT BUFFER
CURSR(IC,IX,IY)	DISPLAY CURSOR ON CRT HALT UNTIL KEYBOARD DEPRESSED IC= KEYBOARD CHARACTER IX= X CURSOR POSITION IY= Y CURSOR POSITION
DSK2(14,NBUF,NWD,JR)	14/15 = READ/WRITE MASTER FILE NBUF= BUFFER NAME NWD= NUMBER OF WORDS JR= RELATIVE SECTOR OF MASTER FILE
ERASE	CLEARs CRT DISPLAY
EXEC(10,NM)	CALL PROGRAM STORED IN DISC FILE NM
EXEC(14,IDCON,NBUF,NWD,NAME,NDX)	
EXEC(15,IDCON,NBUF,NWD,NAME,NDX)	14/15= READ/WRITE DISC FILE IDCON= DISC CONTROL WORD NBUF= START LOCATION OF BUFFER NWD= NUMBER OF WORDS IN BUFFER NAME= NAME OF FILE TO READ/WRITE NDX= STARTING SECTOR OF FILE
JFIND(IST,IEND,IVAL,IFIL)	CONDUCTS BINARY SEARCH FROM IFIL(IST) TO IFIL(IEND) RETURNS INDEX OF LAST VALUE LESS THAN OR EQUAL TO IVAL
JPLOT(K,IX,IY)	PLOT ON CRT K = CONTROL OF POINT OR LINE PLOT IX= X POSITION IY= Y POSITION
POSAC(K)	POSITIONS ALPHANUMERIC CURSOR K= LINE NUMBER
SL2RS(JMFO,NS,JS,JR)	FIND RELATIVE SECTOR FROM MASTER FILE DIRECTORY LOCATION JMFO= MASTER FILE INFORMATION BUFFER NS= NUMBER OF SIGNALS TO FIND JS= DIRECTORY LOCATION JR= RELATIVE SECTOR IN MASTER FILE

INTERFACE PROGRAM

FTN

```

PROGRAM FACE
DIMENSION IFIL(3)
DIMENSION IFL(3)
DIMENSION NM(3),NAME(3)
DIMENSION NAM(3)
COMMON M,MDX,ISIG(126)
COMMON RATE,N,IX(256)
COMMON IC(30)
COMMON IDR(768)
EQUIVALENCE (TIMB,IC(7)),(TT,IC(12))
EQUIVALENCE (TOT,IC(14)),(RST,IC(16))
EQUIVALENCE (DIS,IC(20))

```

THIS PROGRAM REFORMATS THE DISC FILE PARF TO
A FORM ACCEPTABLE BY THE ESL PROGRAM RASF
IT SETS UP A FILE PMAP TO PASS INFORMATION TO THE
RASTER DISPLAY PROGRAM

VARIABLE INFORMATION

```

IFL  NAME OF FILE CONTAINING FORMATTED INFO
NAM  NAME OF FILE TO BE REFORMATTED
NM   NAME OF NEXT PROGRAM
NAME NAME OF CURRENT PROGRAM
RATE,N,IX  BUFFER FOR SIGNAL ENTRY FROM PARF
           RASTER PERIOD,NUMBER OF ENTRIES,
           DELTA TRANSITION TIMES
IC(1)  NUMBER OF WORDS OF FILE INFO
IC(2-4) NAME OF FILE WHERE TIME DATA STORED
IC(5)  # SECTORS IN FILE PMAP
IC(6)  # SIGNALS IN FILE PMAP
IC(7-8) QUANTIZATION INTERVAL
IC(9-11) FILE TYPE CODING
IC(12-13) SIGNAL DURATION
IC(14-15) NUMBER OF TRANSITION TIMES IN FILE
IC(16-17) NOMINAL RASTER PERIOD
IDR    DIRECTORY BUFFER FOR PARF

```

```

DATA IFL(1),IFL(2),IFL(3)
X    /50115B,40520B,20040B/
      PM    AP

```

```

DATA NAM(1),NAM(2),NAM(3)
X    /50101B,51106B,20040B/
      PA    RF

```

```

DATA NM(1),NM(2),NM(3)
X    /51101B,51506B,20040B/
      RA    SF

```

```

DATA NAME(1),NAME(2),NAME(3)
X    /43101B, 41505B,20040B/
      FA    CE

```

```
CALL EXEC (14,102B,IDR,768,NAM,0)
```

```

1  CONTINUE
  READ(5,200) ISG
200 FORMAT(I4)
  IF(ISG) 1,999,201
201 CONTINUE
  DO 10 I=1,768
  IF (IDR(I)-ISG) 10,5,10

```



```

      5 NDX=I*2+4
      GO TO 20
101 10 CONTINUE
C
      WRITE (6,101) ISG
101 101 FORMAT ("SIGNAL NO. ",I5,"NOT PRESENT")
      GO TO 1
C
      20 CONTINUE
      CALL EXEC (14,102B,RATE,256,NAM,NDX)
      CALL EXEC (15,102B,IX,256,IFL,0)
C
      CALCULATE SIGNAL DURATION
C
      TT=0.
      DO 45 I=1,N
      IF (IX(I)) 45,45,44
44 44 CONTINUE
      TT=TT+FLOAT(IX(I))
45 45 CONTINUE
C
      FORMAT FILE INFORMATION
C
      IC(1)=21
      DO 40 I=1,3
      K=I+1
      IC(K)=IFL(I)
40 40 CONTINUE
      IC(5)=2
      IC(6)=1
      IC(9)=52117B
      IC(10)=0
      IC(11)=1002B
      IC(18)=0
      IC(19)=ISG
      RST=-RATE
      TIMB=.0001
      RST=RST*TIMB
      TOT=FLOAT(N)
      DIS=-1.
C
      SET UP FILE PMAP USING ESL ROUTINE
C
      CALL USENT (IC,NAME)
      CALL EXEC (10,NM)
999 999 STOP
      END

```


MEASUREMENT PROGRAM

FTN

```

PROGRAM FOURD
DIMENSION FTX(256)
DIMENSION NAMR(3)
DIMENSION NPAR(3)
COMMON ID(2048)
COMMON BP(256)
COMMON SIG(128)
COMMON IDM(3),SNUM,IDMY(11)
COMMON NAME(3)
COMMON TT,T1,T2
COMMON RATE,PER,TB,MODE
COMMON IDMM(9)
COMMON JX,JY
COMMON MOD1
COMMON ISG
EQUIVALENCE (FTX(1),ID(1025))
EQUIVALENCE (NDX,BP(256))

```

C
C
C
C
C
C
C
C
C
C

THIS PROGRAM IS A DRIVING ROUTINE WHICH
CALLS THE MEASUREMENT AND INTERPOLATION
SUBROUTINE ,FAST FOURIER TRANSFORM,AND
THE FOURIER SPECTRUM DISPLAY

THE FOLLOWING VARIABLES ARE USED:

FTX COMPLEX INPUT TO FAST FOURIER TRANSFORM
SIG MEASURED SIGNAL PARAMETERS

C
C
C
C
C
C
C
C

```

DATA NAMR(1),NAMR(2),NAMR(3)
X /43101B,41505B,20040B/
FA CE
DATA NPAR(1),NPAR(2),NPAR(3)
X /50101B,51040B,20040B/
PA R

```

ZERO PARAMETER MEASUREMENTS

```

DO 5 I=1,128
5 SIG(I) = 0.
CALL RESDU

```

C
C
C
C

PAUSE FOR RASTER DISPLAY
PRESS Q TO QUIT

```

CALL CURSR(IC,IX,IY)
IF(IC-121B) 10,20,10
10 CONTINUE

```

C
C
C

CALCULATE 128 COMPLEX POINT FAST FOURIER TRANSFORM

```

CALL FOUR2(FTX,128,1,-1)
CONTINUE
CALL FORD

```

C
C
C

PAUSE FOR FOURIER DISPLAY

```

CALL CURSR(IC,IX,IY)
IF(IC-121B) 30,20,30

```

C
C
C
C
C

WRITE PARAMETERS TO DISC FILE PAR

```

30 CALL EXEC(15,1028,SIG,256,NPAR,NDX)

```


C CALL PROGRAM FACE FROM DISC

C
20 CONTINUE
CALL EXEC (10,NAMR)
END\$

SUBROUTINE RESDU
DIMENSION IT(256)
DIMENSION TX(2)
DIMENSION FTX(256),XF(2,4),B(5)
DIMENSION X(2,128)
COMMON ID(2048)
COMMON BP(256)
COMMON SIG(128)
COMMON IDM(3),SNUM,IDMY(11)
COMMON NAME(3)
COMMON TT,T1,T2
COMMON RATE,PER,TB,MODE
COMMON LCT,SCX,SCR,STX,STR
COMMON JX,JY
COMMON MOD1
COMMON ISG
EQUIVALENCE (X(1,1),ID(1537)),(FTX(1),ID(1025))
EQUIVALENCE(IT,ID(513))
EQUIVALENCE(TX,T1)
POLY(F)=(SIG(5)+F*(SIG(6)+F*(SIG(7)
1+F*(SIG(8)+F*SIG(9))))/PER

C
C THIS SUBROUTINE INTERPOLATES BETWEEN PTS OF THE
C SCAN, PROVIDES MEAN SQUARE FITS FOR EACH BIAS
C AND OBTAINS A RASTER VARIANCE MEASURE
C

C THE FOLLOWING VARIABLES APPEAR IN COMMON
C

C ID(1-256) DELTA TRANSITION TIMES
C IT TRANSITION TIMES
C X(1,K) RASTER PHASE POINTS
C X(2,K) FLOATING PT TRANSITION TIME
C SIG SIGNAL PARAMETERS MEASURED
C SNUM TOTAL NUMBER OF TRANSITIONS
C TT SIGNAL DURATION
C RATE,PER,TB,MODE,LCT,SCX,SCR,STX,STR,JX,JY
C CONTROL VARIABLES FOR DISPLAY
C ISG SIGNAL IDENTIFIER
C XF(2,4) 4 POINTS USED FOR CUBIC INTERPOLATION
C B COEFFICIENTS OF INTERPOLATED FIT
C

C INITIALIZATION
C

NTOP=IFIX(SNUM)
LT=0
RX=JX
RY=JY
LX=900
LY=500
SCAX=FLOAT(LX)/TT
RAT2=PER/2.
RAT4=RAT2/2.

C CONVERT DELTA TIME TO TIME
C

DO 200 I=1,NTOP
IF(ID(I)) 200,200,201
201 IDEL=ID(I)
LT=LT+IDEL
200 IT(I)=LT

C FIND STARTING INDEX FOR MINIMAL PHASE DIFFERENCE
C

JST=JFIND(1,NTOP,400,IT)


```

    SIG(2)=FLOAT(IT(NTOP-2)-400)
    DELT= SIG(2)/128.
    SIG(2)=1./(SIG(2)*TB)
    NDX=1
    NSWCH=1
    OHSQ=0.
    JTOP=0
    KST=1
999  CONTINUE
    PHM=0.
    PHSQ=0.

C
C
C    INITIALIZE RASTER PHASE POINT
    J=JST
210  J=J-2
    IF(J-1) 212,212,209
212  PHH=0.
    TST=0.
    GO TO 214
209  IF(ID(J)) 210,210,213
213  TST=IT(J)
    PHH=AMOD(TST,PER)
    IF(PHH-RAT2) 214,215
215  PHH=PHH-PER
214  X(1,1)=TST
    X(2,1)=PHH
    DELL=0.
    PHM=PHM+PHH
    PHSQ=PHSQ+PHH*PHH
    N=TST/PER
    J=1
    PHL=PHH

C
C
C    CALCULATE MINIMAL PHASE DIFFERENCE POINTS
    DO 300 I=JST,NTOP,2
    IF(ID(I)) 300,300,302
302  TST=IT(I)
    41  N=N+1
    PHH=TST-FLOAT(N)*PER
    DEL=PHH-PHL
    IF(ABS(DEL)-RAT2) 42,42,41
    42  N=N-1

C
C
C    UPDATE MEANS AND VARIANCE ESTIMATES
    PHM=PHM+PHH
    PHSQ=PHSQ+PHH*PHH
    OHSQ=OHSQ+DEL*DEL
    IF(ABS(DEL)-RAT4) 305,707
707  DEL=DELL*.5
    PHH=PHL+DEL
305  J=J+1
    X(1,J)=TST
    X(2,J)=PHH
    DELL=DEL
    PHL=PHH
300  CONTINUE

C
C
C    CALCULATE MEAN AND VARIANCE FOR BIAS CONDITION
    JTOP=JTOP+J
    JP=J
    XNUM=JP
    XNUM1= JP-1
    SIG(35) = SIG(35)+(PHSQ-PHM*PHM/XNUM)/XNUM1
    PHM1=PHM/XNUM

C
C
C    CALCULATE POLYNOMIAL MEAN SQUARE FIT

```



```

C      CALL CURV(JP)
C
C      INTERPOLATE BETWEEN POINTS
C
      K=1
      CALL CURF(X(1,K),3,B)
      XP=400.
      DO 601 I=1,128
607    IF(XP-X(1,K+2)) 602,603
602    CONTINUE
      DX=XP-X(1,K)
      FTX(KST)=(B(1)+DX*(B(2)+DX*(B(3)+DX*B(4))))/PER
      IX=XP*SCAX+RX
      IY=(FTX(KST)+.5)*FLOAT(LY)+RY
C
C      DISPLAY INTERPOLATION
C
      CALL JPLOT(0,IX-4,IY-6)
      CALL ALPHA
      CALL CHOUT(52B)
C
C      INPUT FOR FAST FOURIER TRANSFORM
C
      FTX(KST)=FTX(KST)-POLY(XP)
      KST=KST+2
      XP=XP+DELT
      GO TO 601
603    IF(K+4-JP) 604,602,602
604    CONTINUE
      K=K+1
      CALL CURF(X(1,K),3,B)
      GO TO 607
601    CONTINUE
      CALL CHOUT(-1)
C
C      PLOT POLYNOMIAL MEAN SQUARE FIT DATA
C
      XP=400.
      DO 20 I=1,128
      IX=XP*SCAX+RX
      IY=(POLY(XP)+.5)*FLOAT(LY)+RY
      CALL JPLOT(0,IX-4,IY-6)
      CALL ALPHA
      CALL CHOUT(53B)
      XP=XP+DELT
20    CONTINUE
      CALL CHOUT(-1)
C
C      CHECK TO SEE IF BOTH BIAS CONDITIONS ARE PLOTTED
C
      NSWCH=-NSWCH
      IF(NSWCH) 700,701
700    CONTINUE
      XP=400.
      PHM2=PHM1
      JST=JST+1
      KST=2
      SIG(4)=SIG(3)
      DO 702 I=1,5
      K1=I+4
      K2=I+9
      SIG(K2)=SIG(K1)
702    CONTINUE
      GO TO 999
701    CONTINUE
C
C      CALCULATE MEAN BIAS AND INTRINSIC VARIANCE
C
      XNM=JTOP*2
      SIG(20)=ABS(PHM2-PHM1)
      SIG(36)=OHSQ/XNM

```


C
C
C
C

CALCULATE LAGUERRE COEFFICIENTS

```
P=80.
TEND=800.
CALL LAGER(ID,PER,P,TEND,SIG(15))
RETURN
END
```

```
SUBROUTINE CURV(NPTS)
DIMENSION X(2,128)
DIMENSION B(5)
DIMENSION A(5,6)
COMMON ID(2048)
COMMON BP(256)
COMMON SIG(128)
COMMON IDMMY(16)
COMMON NAME(3)
COMMON TT,T1,T2
COMMON RATE,PER,TB,MODE
COMMON LCT,SCX,SCR,STX,STR
COMMON JX,JY
COMMON MOD1,ISG
EQUIVALENCE (X(1,1),ID(1537)),(B(1),SIG(5))
```

C
C
C
C
C
C
C
C

THIS SUBROUTINE CALCULATES THE COEFFICIENTS
OF A POLYNOMIAL MEAN SQUARE FIT OF ORDER LESS
THAN FOUR, USING UP TO 128 SAMPLE POINTS

X(2,128) SAMPLE POINTS
SIG(5-10) POLYNOMIAL COEFFICIENTS

SPECIFY ORDER OF POLYNOMIAL TO BE FIT

```
NUM=2
DO 20 J=1,6
DO 20 I=1,5
A(I,J)= 0.
20 CONTINUE
A(1,2)=FLOAT(NPTS)
DO 100 I=1,NPTS
YT= X(2,I)
X1=X(1,I)
X2= X1*X1
X3= X2*X1
X4= X3*X1
X5= X4*X1
X6= X5*X1
X7= X6*X1
X8= X7*X1
C
A(1,1)= A(1,1)+YT
A(1,3)= A(1,3)+X1
A(1,4)=A(1,4)+X2
A(1,5)=A(1,5)+X3
A(1,6)=A(1,6)+X4
A(2,1)=A(2,1)+X1*YT
A(2,6)=A(2,6)+X5
A(3,1)=A(3,1)+X2*YT
A(3,6)=A(3,6)+X6
A(4,1)=A(4,1)+X3*YT
A(4,6)=A(4,6)+X7
A(5,1)=A(5,1)+X4*YT
A(5,6)=A(5,6)+X8
100 CONTINUE
DO 25 I=2,5
DO 25 J=2,5
A(I,J)=A(I-1,J+1)
```

C


```

25 CONTINUE
   LCT= NUM+2
   LMT= NUM+1
   DO 35 I=1,NUM
     DO 35 J=1,I
       L= J+1
       K=I+1-J
       M=K+1
       FAC= A(K,L)/A(M,L)
       CALL ROW(A,NUM,M,FAC,K)
35  CONTINUE
     M= LCT
     DO 45 I=1,LMT
       DO 36 K=1,LCT
         A(I,K)=A(I,K)/A(I,M)
36  CONTINUE
       IF(I-LMT) 37,45,45
37  CONTINUE
       DO 38 J=I,NUM
         MK=J+1
         FAC=A(MK,M)
         CALL ROW(A,NUM,I,FAC,MK)
38  CONTINUE
       M= M-1
45  CONTINUE
       J=LCT
       DO 46 I=1,LMT
         J=J-1
         B(J)= A(I,1)
46  CONTINUE
       IF(LCT-5) 70,70,75
70  DO 47 I=LCT,5
         B(I)= 0.
47  CONTINUE
75  RETURN
    END

```

C
C

```

SUBROUTINE ROW (A,NUM,IRM,FAC,IRS)
  DIMENSION A(5,6)
  DIMENSION TEMP(6)
  KJ=NUM+2
  DO 50 I=1,KJ
    TEMP(I)= FAC*A(IRM,I)
50  CONTINUE
    DO 100 I=1,KJ
      A(IRS,I)= A(IRS,I)-TEMP(I)
100 CONTINUE
  RETURN
  END
  END$

```

```

SUBROUTINE LAGER(ID,RATE,P,TEND SL)
  DIMENSION ID(256),SL(5)

```

C
C
C
C
C

THIS SUBROUTINE CALCULATES THE LAGUERRE POLYNOMIAL
COEFFICIENTS FOR THE FIRST FIVE POLYNOMIALS USING
TRAPEZOIDAL INTEGRATION

```

DO 5 I= 1,5
5  SL(I)=0.

```

C
C
C
C

CALCULATE MINIMAL PHASE DIFFERENCE RASTER POINTS
USING BOTH BIAS CONDITIONS

```

PHL=0.
TT= 0.
N= 0
P= 1./P

```



```

RAT2 =RATE/2.
DO 10 I=1,256
IF(ID(I)) 10,10,11
11 TT=TT+FLOAT(ID(I))
IF(TT-TEND) 12,12,13
12 DEL= FLOAT(ID(I))
14 N=N+1
PHH=(TT-FLOAT(N)*RATE)
DIF= PHH-PHL
IF(ABS(DIF)-RAT2) 15,15,14
15 N=N-1

```

C
C
C

```

CALCULATE CONTRIBUTIONS TO LAGUERRE POLYNOMIALS

K=1
T1=TT-DEL
30 K=K-1
DIF=PHL
RT=P*T1
REXP=.5*DEL*EXP(-RT)
SL(1)= SL(1)+REXP*DIF
SL(2)=SL(2)+(2.*RT-1.)*REXP*DIF
SL(3)=SL(3)+(1.+RT*(-4.+RT*(2.)))*REXP*DIF
SL(4)=SL(4)+(-1.+RT*(6.+RT*(-6.+RT*(4./3.)))*REXP*DIF
SL(5)=SL(5)+(1.+RT*(-8.+RT*(12.+RT*(-16./3.+RT*2./3.))
1 ))*REXP*DIF
1 *REXP*DIF
T1=TT
PHL=PHH
IF(K) 31,30,30
31 CONTINUE
10 CONTINUE

```

C
C
C

```

NORMALIZE
13 PNORM=SQRT(2./P)
DO 20 I=1,5
20 SL(I)=SL(I)*PNORM
RETURN
END

```

```

SUBROUTINE FORD
DIMENSION FX(256)
DIMENSION NM(3)
DIMENSION FTX(256)
COMMON ID(2048)
COMMON BP(256)
COMMON SIG(128)
COMMON IDM(3),SNUM,IDMY(11)
COMMON NAME(3)
COMMON TT,T1,T2
COMMON RATE,PER,TB,MODE
COMMON IDMM(9)
COMMON JX,JY
COMMON MOD1
COMMON ISG
EQUIVALENCE (FX(1),ID(1537))
EQUIVALENCE (FTX(1),ID(1025))
EQUIVALENCE (NDX,BP(256))

```

C
C
C
C

THIS SUBROUTINE CALCULATES AND PLOTS FOURIER
MAGNITUDES FOR THE 64 TERMS

```

RX=JX
RY=JY
LX= 900
LY= 500
XMAX= 200.
YMAX=4.

```



```

      XMX4= XMAX/4.
      YMX2= YMAX/2.
      SIG(1)= RATE
      STX=0.
      STY=0.
      SCX= FLOAT(LX)/XMAX
      SCY= FLOAT(LY)/YMAX
      CALL ERASE
      CALL POSAC(2)
      WRITE(6,100) ISG,SIG(37),SIG(2)
100  FORMAT("FREQ MAG PLOT FOR SIG NO.",I5/
1    "VARIANCE OF RESIDUE=",F10.6.
2    " RESOLUTION= ",F10.6)
      CALL AXIS(JX,JY,SCX,SCY,STX,STY,
1        XMX4,YMX2,XMAX,YMAX)

C
C
      CALCULATE FIRST 64 MAGNITUDE COMPONENTS

      DO 40 I=1,63
      J=2*I+1
      K=257-2*I
      L=I+1
      JI=J+1
      KI=K+1
      FX(L)=(SQRT(FTX(J)**2+FTX(K)**2+FTX(JI)**2
1    +FTX(KI)**2))/2.
40  CONTINUE
      FX(1)=(SQRT(2.*(FTX(1)**2+FTX(2)**2)))/2.

C
C
      PLOT SPECTRUM

      DO 50 I=1,64
      IX=FLOAT(I-1)*SIG(2)*SCX+RX
      IY=FX(I)*SCY+RY
      CALL JPLOT(1-I,IX,IY)
50  CONTINUE
      CALL ALPHA
      CALL CHOUT(-1)

C
C
      DO 10 I=1,64
      K=I+39
      SIG(K)=FX(I)
10  CONTINUE
      RETURN
      END

```


STATISTICS AND SCALING PROGRAM

FTN

```

PROGRAM CSTAT
DIMENSION SPAR(50),JW(50)
DIMENSION NPAR(3),NCLS(3),NLIB(3)
COMMON JSAVE(50)
COMMON JMFO(128),JMFD(768),MFE(384)
COMMON IPD(256),SIG(128)
COMMON ICD(128),ISG(128)
COMMON JBUF(4),CLMN(50),CLVA(50),JNPA(50),JPNUM(50),
1 JCDO(80)
COMMON JGLOB(4),GM(50),GV(50),GTV(50),JBND(50),
1 JGDM(30)
EQUIVALENCE (SPAR(1),MFE(5)),(JW(1),MFE(105))
EQUIVALENCE (IDNO,JBUF(1)),(NENT,JBUF(3))

THIS PROGRAM PLACES MEASURED SIGNAL PARAMETERS IN THE
SIGNAL FILES OF THE MASTER FILE. CLASS STATISTICS
ARE CALCULATED AND STORED IN THE MASTER FILE CLASS
ENTRIES.

THE FOLLOWING VARIABLES ARE USED:
JSAVE STORAGE FOR OLD PARAMETERS
JMFO INFORMATION ON MASTER FILE STRUCTURE
JMFD MASTER FILE DIRECTORY
MFE BUFFER FOR MASTER FILE ENTRY
IPD PAR FILE DIRECTORY
SIG BUFFER FOR PAR FILE ENTRY
ICD CSIGF FILE DIRECTORY
ISG BUFFER FOR SET OF SIGNAL NUMBERS OF SAME
CLASS
JBUF-JCDO BUFFER FOR ENTRY TO CLASS FILE
JGLOB-JGDM BUFFER FOR ENTRY TO GLOBAL FILE

DATA NPAR(1),NPAR(2),NPAR(3)
1 /50101B, 51040B, 20040B/
PA R
DATA NCLS(1),NCLS(2),NCLS(3)
1 /41523B, 44507B, 43040B/
CS IG F
DATA NLIB(1),NLIB(2),NLIB(3)
1 /46501B,51524B,43040B/
MA ST F

READ IN DIRECTORIES FOR FILE PAR,CSIGF,MASTF

CALL EXEC(14,102B,IPD,256,NPAR,0)
CALL EXEC(14,102B,ICD,128,NCLS,0)
CALL EXEC(14,102B,JMFO,128,NLIB,0)
NWD=JMFO(2)
CALL DSK2(14,JMFD,NWD,1)
CALL SL2RS(JMFO,1,1,LG)
CALL DSK2(14,JGLOB,384,LG)

ZERO GLOBAL MEANS AND STND DEV

DO 10 I=1,50
GM(I)=0.
GV(I)=0.
GTV(I)=0.
10 CONTINUE
JGLOB(3)=0
CALL ERASE

```



```

      CALL POSAC(0)
      WRITE(6,113)
113  FORMAT("ENTER INPUT DEVICE (5 FOR CR, 6 FOR TEK) ")
C
C      READ(1,*) LU
      READ(LU,*) N
C
C      CHANGE PARAMETER NAMES IN MASTER FILE INFO SECTOR
C
      DO 90 I1=1,N
      K1= I1+14
      JSAVE(I1)=JMFO(K1)
      READ(LU,*) JMFO(K1)
90  CONTINUE
      JMFO(14)=N
      1 CONTINUE
      CALL ERASE
      CALL POSAC(0)
      WRITE(6,100)
100  FORMAT("ENTER CLASS NO."/)
      READ(6,*) NUM
C
C      READ IN SIGNAL IN CLASS SPECIFIED
C
      CALL EXEC(14,102B,ISG,128,NCLS,NUM)
      IDNO=ICD(NUM)
C
C      FIND CLASS ENTRY IN MASTF AND READ INTO JBUF
C
      DO 70 JM=1,NWD
      IF (JMFO(JM)+IDNO) 70,71,70
70  CONTINUE
      GO TO 1
71  CONTINUE
      CALL SL2RS(JMFO,1,JM,JR)
      LOCLI=JR
      CALL DSK2(14,JBUF,384,LOCLI)
C
C      ZERO FILE MEANS AND VARIANCE
C
      DO 11 I=1,50
      JNPA(I)=0
      CLMN(I)=0.
      CLVA(I)=0.
11  CONTINUE
      NENT=0
C
C      COMPUTE MEAN + VARIANCE OF PARAMETERS.
C
      DO 20 K=1,128
      IF(ISG(K)) 21,21,22
C
C      FIND RELATIVE SECTOR OF ISG
C
      22 DO 30 J=1,256
      IF(IPD(J)-ISG(K)) 30,23,30
      30 CONTINUE
      WRITE(6,101) ISG(K)
101  FORMAT("SIGNAL NO.",I6," NOT FOUND"/)
      GO TO 20
      23 NENT=NENT+1
      JGLOB(3)=JGLOB(3)+1
      NDX=2*J+2
      CALL EXEC(14, 102B,SIG,256,NPAR,NDX)
      CALL TRNSG(ISG(K),IDNO)
C
C      CALCULATE SUM + SUM OF SQUARES
C
      DO 40 I=1,50
      A=SPAR(I)
      GM(I)=GM(I)+A

```



```

    GV(I)=GV(I)+A*A
    IF(JW(I)) 40,40,50
50 CONTINUE
    JNPA(I)=JNPA(I)+1
    XNUM=FLOAT(JNPA(I))
    XNUM1=XNUM-1.
    XNUM2=XNUM-2.
    CLMN(I)=(CLMN(I)*XNUM1+SPAR(I))/XNUM
    IF(JNPA(I)-1) 51,51,52
52 CONTINUE
    CLVA(I)=XNUM2/XNUM1*CLVA(I)+((SPAR(I)-CLMN(I))*2)/
1      XNUM
51 CONTINUE
40 CONTINUE
20 CONTINUE
21 CONTINUE
    CALL CURSR(IC,IX,IY)
    DO 80 I=1,50
        JM=I+14
        JPNUM(I)=JMFO(JM)
80 CONTINUE
    CALL DSK2(15,JBUF,384,LOCLI)
    CALL ERASE
    CALL POSAC(0)
    WRITE(6,102) IDNO
102 FORMAT("CLASS NAME      ",I6/
1"PAR      MEAN      VARIANCE      #USED"/)
    WRITE(6,103) (I,CLMN(I),CLVA(I),JNPA(I),I=1,50)
103 FORMAT(2(I6,2E12.4,I4,2X))
    CALL CURSR(IC,IX,IY)
    IF(IC-121B) 1,999,1
999 CONTINUE

```

C
C
C

CALCULATE GLOBAL MEAN AND STD DEV

```

    CALL ERASE
    CALL POSAC(0)
    WRITE(6,800)
800 FORMAT("GLOBAL VALUES"/,
1"PAR      GLOBAL MEAN      STD DEV      TYP DEV      JBND")
801 FORMAT(6X,I4,3(2X,E12.4),1X,I4)
    C=JGLOB(3)
    DO 802 I=1,50
        A=GM(I)
        B=GV(I)
        A=A/C
        D=SQRT((B-C*A*A)/(C-1.))
        IF(D) 803,803,804
803 D=1.
804 GM(I)=A
        GV(I)=D
        GTV(I)=.2*D
        JBND(I)=2
802 CONTINUE

```

C
C
C

SET UP VALUES IN SECTOR 0 OF MASTER FILE

```

    JMFO(5)=-1
    JMFO(6)=5
    CALL EXEC(15,102B,JMFO,128,NLIB,0)

```

C
C
C

PRINT OUT VALUES OF GLOBAL FILE

```

    WRITE(6,801)(JMFO(I+14),GM(I),GV(I),GTV(I),
1JBND(I),I=1,50)
    CALL DSK2(15,JGLOB,384,LG)
    STOP
    END

```



```

SUBROUTINE TRNSG(IS,LDUM)
DIMENSION WSC(15)
DIMENSION SPAR(50),JW(50)
DIMENSION NPAR(3),NCLS(3),NLIB(3)
COMMON JSAVE(50)
COMMON JMFO(128),JMFD(768),MFE(384)
COMMON IPD(256),SIG(128)
COMMON ICD(128),ISG(128)
COMMON JBUF(4),CLMN(50),CLVA(50),JNPA(50),JPNUM(50),
1 JCDO(80)
COMMON JGLOB(4),GM(50),GV(50),GTV(50),JBND(50),
1 JGDM(30)
EQUIVALENCE (SPAR(1),MFE(5)),(JW(1),MFE(105))
EQUIVALENCE (IDNO,JBUF(1)),(NENT,JBUF(3))

```

THIS SUBROUTINE CALCULATES PARAMETERS AND STORES THEM
IN THE MASTER FILE BUFFER MFE

```

DATA WSC(1),WSC(2),WSC(3),WSC(4)
1 /1000.,1.,1000.,1000./
DATA WSC(5),WSC(6),WSC(7),WSC(8),WSC(9)
1 /1.,100.,1.E06,1.E09,1.E13/
DATA WSC(10),WSC(11),WSC(12),WSC(13),WSC(14)
1 /1.,100.,1.E06,1.E09,1.E13/

```

THIS SUBROUTINE SCALES PARAMETERS AND STORES THEM
IN THE MASTER FILE BUFFER MFE

```

NTOP=JMFO(2)
DO 5 I=1,NTOP
IF (IS-JMFD(I)) 5,6,5
5 CONTINUE
DO 7 I=1,50
JW(I)=0
7 CONTINUE
WRITE(6,100)
100 FORMAT(15,1X,"NOT IN MASTER")
RETURN
6 CONTINUE
K=0
CALL SL2RS(JMFO,1,I,JR)
CALL DSK2(14,MFE,384,JR)
MFE(1)=IS
MFE(3)=-JBUF(1)
MFE(4)=0
K=K+1
JWT=1

```

RATE

SPAR(1)= SIG(1)*WSC(1)

RESOLUTION

SPAR(2)= SIG(2)

POLYNOMIAL MEAN SQUARE FIT COEFFICIENTS

```

SPAR(3)= SIG(5)*WSC(5)
SPAR(4)=SIG(6)*WSC(6)
SPAR(5)= SIG(7)*WSC(7)
SPAR(6)= SIG(10)*WSC(10)
SPAR(7)=SIG(11)*WSC(11)
SPAR(8)= SIG(12)*WSC(12)

```

LAGUERRE POLYNOMIAL COEFFICIENTS


```

      SPAR(9)= SIG(15)
      SPAR(10)= SIG(16)
      SPAR(11)= SIG(17)
      SPAR(12)=SIG(18)
      SPAR(13)= SIG(19)
C
C      MEAN BIAS
C
      SPAR(14)= SIG(20)
C
C      VARIANCE
C
      SPAR(15)= SIG(35)
C
C      INTRINSIC VARIANCE
C
      SPAR(16)= SIG(36)
C
C      NONPARAMETRIC CORRELATION
C
      SPAR(17)= 1.-SIG(36)/SIG(35)
C
C      FOURIER POWER SPECTRUM USING LINEAR INTERPOLATION
C
      K= 18
      J=40
      L=1
      RES=SIG(2)
      SPAR(K)=ABS(SIG(J))
      DO 60 I=1,32
      K=K+1
63  CONTINUE
      DIFF=FLOAT(I)*2.-FLOAT(L)*RES
      IF(DIFF) 61,61,62
61  SPAR(K)=SIG(J+1)+(SIG(J+1)-SIG(J))*DIFF/RES
      GO TO 60
62  L=L+1
      J=J+1
      GO TO 63
60  CONTINUE
256 CONTINUE
      DO 70 I=1,50
      JW(I)=1
70  CONTINUE
      CALL DSK2(15,MFE,384,JR)
      RETURN
      END

```


LIST OF REFERENCES

1. Aizerman, M.A., "Remarks on Two Problems Connected with Pattern Recognition," Methodologies of Pattern Recognition, p. 1-10, S. Watanabe, Editor, Academic Press, 1969.
2. Andrews, H.C., Introduction to Mathematical Techniques in Pattern Recognition, Wiley-Interscience, 1972.
3. Chang, Chieng-Yi, "Dynamic Programming as Applied to Feature Selection in a Pattern Recognition System," IEEE Transactions on Systems Management Cybernetics, v. SMC-3, n. 2, p. 166-173, March, 1973.
4. Cooper, D.B., "When Should a Learning Machine Ask for Help?" IEEE Transactions on Information Theory, v. IT-20, no. 4, p. 455-471, July 1974.
5. Cover, T.M., "Estimation by the Nearest Neighbor Rule," IEEE Transactions on Information Theory, v. IT-14, no. 1, p. 50-55, January 1968.
6. Cover, T.M. and Hart, P.E., "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, v. IT-13, p. 21-27, January, 1967.
7. Duda, R.O. and Hart, P.E., Pattern Recognition and Scheme Analysis, Wiley-Interscience, 1973.
8. Fralick, S.C. and Scott, R.W., "Nonparametric Bayes-Risk Estimation," IEEE Transactions on Information Theory, v. IT-17, no. 4, p. 440-444, July, 1971.
9. Fu, K.S., Sequential Methods in Pattern Recognition and Machine Learning, Academic Press, 1968.
10. Gold, B. and Rader, C.M., Digital Processing of Signals, McGraw-Hill, p. 159-202, 1969.
11. Hughes, G.F., "On the Mean Accuracy of Statistical Pattern Recognizers," IEEE Transactions on Information Theory, v. IT-14, no. 1, p. 55-63, January, 1968.
12. Jarvis, R.A. and Patrick, E.A., "Clustering Using a Similarity Measure Based on Shared Near Neighbors," IEEE Transactions on Computers, v. C-22, no. 11, p. 1025-1034, November, 1973.

13. The Johns Hopkins University, Representation and Analysis of Signals, Part XXIV, Statistical Estimation of Intrinsic Dimensionality and Parameter Identification, by G.V. Trunk, June 1967.
14. Meisel, W.S., Computer-Oriented Approaches to Pattern Recognition, Academic Press, 1972.
15. Meloy, J.D., Feasibility of Pulse Signal Classification by Spectral Parameters, Electrical Engineer Thesis, Naval Postgraduate School, Monterey, California, December, 1973.
16. Michael, M.T. and Lin, W.C., "Experimental Study of Information Measure and Inter-Intra Class Distance Ratios on Feature Selection and Orderings," IEEE Transactions on Systems Management and Cybernetics, v. SMC-3, no. 2, p. 172-181, March 1973.
17. Mucciardi, A.N. and Gose, E.E., "A Comparison of Seven Techniques for Choosing Pattern Recognition Properties," IEEE Transactions on Computers, v. C-20, no. 9, p. 623-1031, September 1971.
18. Patrick, E.A., Fundamentals of Pattern Recognition, Prentice-Hall, 1972.
19. Purdue University, School of Electrical Engineering, TR-EE-69-18, Unsupervised Estimation and Processing of Unknown Signal, by E.A. Patrick and J.P. Costello, July 1969.
20. Purdue University, School of Electrical Engineering, TR-EE-69-23, Computer Analysis and Classification of Waveforms and Pictures, Part I, Waveforms, by E.A. Patrick, F.P. Fischer, and L.Y.L. Shen, July 1969.
21. Toussaint, G.T., "Bibliography on Estimation of Misclassification," IEEE Transactions on Information Theory, v. IT-20, no. 4, p. 472-478, July 1974.
22. Willsky, A.S., "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication, Part I," IEEE Transactions on Information Theory, v. IT-20, no. 5, p. 577-584, September 1974.

INITIAL DISTRIBUTION LIST

	No. Copies
1. Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2. Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3. Assoc. Professor Stephen Jauregui, Jr. Code 52Ja Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	10
4. National Security Agency W. Group Fort G.G. Meade, Maryland 20755 ATTN: James Boone	1
5. Naval Security Group Headquarters 3801 Nebraska Avenue, N.W. Washington, D.C. 20390 ATTN: CDR H. Orejuela	3
6. Naval Electronic Systems Command PME 107 Washington, D.C. 20360 ATTN: R. Shields	2
7. LT Thomas A. Reglein USN 2219 Jones Street Sioux City, Iowa 51104	1
8. Electromagnetic Systems Laboratory 495 Java Drive Sunnyvale, California 94086 ATTN: Bill Phillips	1
9. Bunker Ramo Incorporated 31717 La Tienda Drive Westlake Village, California 91361 ATTN: Harry Foulkrod	1
10. Naval Electronics Laboratory Center San Diego, California 92152 ATTN: John Griffin	1

Thesis

160074

R293 Reglein

c.1

Feature selection for
the identification of
bauded signals.

4

or

Thesis

160074

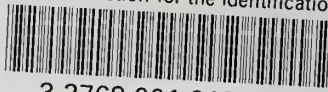
R293 Reglein

c.1

Feature selection for
the identification of
bauded signals.

thesR293

Feature selection for the identification



3 2768 001 01305 5

DUDLEY KNOX LIBRARY